

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 15.6054 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810
Sequence: 1 TDRVSGNLRIRARVLYDFVN.....KARAEEKPAPSDRAGDDAAR 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTCUS-COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3810	100.0	741	US-09-001-984C-106	Sequence 106, App
2	3810	100.0	741	US-09-396-347F-106	Sequence 106, App
3	2570	67.5	731	US-09-252-991A-22524	Sequence 22524, A
4	2289	60.1	739	US-08-836-943-2	Sequence 2, Appli
5	2225	58.4	729	US-09-328-352-7603	Sequence 7603, Ap
6	139	3.6	3892	US-09-328-352-5503	Sequence 5503, Ap
7	135	3.5	3816	US-09-428-517-3	Sequence 3, Appli
8	128	3.4	878	US-08-237-919-2	Sequence 2, Appli
9	128	3.4	878	US-08-732-429-2	Sequence 2, Appli
10	128	3.4	878	US-09-798-267-2	Sequence 2, Appli
11	128	3.4	878	US-09-798-267-3	Sequence 3, Appli
12	128	3.4	878	PCT-US95-05518-2	Sequence 2, Appli
13	128	3.4	11877	US-09-105-537-6	Sequence 6, Appli
14	127	3.3	1850	US-09-252-931A-21798	Sequence 21798, A
15	126.5	3.3	3739	US-09-320-878-2	Sequence 2, Appli
16	126.5	3.3	3739	US-09-105-537-33	Sequence 33, Appli
17	126.5	3.3	3739	US-09-141-908-3	Sequence 3, Appli
18	126.5	3.3	3739	US-09-657-440-2	Sequence 2, Appli
19	125.5	3.3	2039	US-09-077-098A-7	Sequence 7, Appli
20	124	3.3	1346	US-09-105-537-37	Sequence 37, Appli
21	124	3.3	15281	US-08-471-119A-2	Sequence 2, Appli
22	123	3.2	1346	US-09-320-878-4	Sequence 4, Appli
23	123	3.2	1346	US-09-141-908-5	Sequence 5, Appli
24	123	3.2	1346	US-09-657-440-4	Sequence 4, Appli
25	123	3.2	1529	US-09-134-001C-3945	Sequence 3945, Ap
26	120.5	3.2	675	US-09-252-991A-27026	Sequence 27026, A
27	119.5	3.1	2032	US-09-071-035-458	Sequence 458, App

28	119.5	3.1	2032	4	US-09-071-035-462	Sequence 462, App
29	119.5	3.1	2032	4	US-09-071-035-466	Sequence 466, App
30	119.5	3.1	7257	3	US-08-335-409-5	Sequence 5, Appli
31	119.5	3.1	7257	4	US-09-568-102-5	Sequence 5, Appli
32	119.5	3.1	7257	4	US-09-567-969-5	Sequence 5, Appli
33	119.5	3.1	7257	4	US-09-568-480-5	Sequence 5, Appli
34	119.5	3.1	7257	4	US-09-568-486-5	Sequence 5, Appli
35	119.5	3.1	7257	4	US-09-568-472-5	Sequence 5, Appli
36	119.5	3.1	7257	4	US-09-567-899-5	Sequence 5, Appli
37	118.5	3.1	915	4	US-09-206-942-35	Sequence 35, Appli
38	118.5	3.1	1222	4	US-09-206-942-37	Sequence 37, Appli
39	118.5	3.1	1222	4	US-09-206-942-34	Sequence 34, Appli
40	118	3.1	1385	4	US-09-252-991A-21919	Sequence 21919, A
41	116.5	3.1	1289	2	US-08-853-659A-51	Sequence 51, Appli
42	116.5	3.1	4928	3	US-09-036-987A-5	Sequence 5, Appli
43	116.5	3.1	4928	3	US-09-370-700-5	Sequence 5, Appli
44	116.5	3.1	4928	4	US-09-603-207-5	Sequence 5, Appli
45	115	3.0	2042	4	US-09-077-098A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Belisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001.984C
; CURRENT FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034.003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-001-984C-106

Query Match	100.0%;	Score 3810;	DB 3;	Length 741;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	740;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	1	TDRVSGNLRIRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ	60	
Db	2	TDRVSGNLRIRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ	61	
Qy	61	AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV	120	
Db	62	AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV	121	
Qy	121	LNARFALNANARWSLYDALYCTDVIPTDGAKEGPTYNKVGDKVIARXKFLDSDVP	180	
Db	122	LNARFALNANARWSLYDALYCTDVIPTDGAKEGPTYNKVGDKVIARXKFLDSDVP	181	
Qy	181	LSSGSFGDATGFTVQDQQLVVALPKSTGLANPQAGYTGAAESPTSVLLINRGLHIEI	240	
Db	182	LSSGSFGDATGFTVQDQQLVVALPKSTGLANPQAGYTGAAESPTSVLLINRGLHIEI	241	
Qy	241	LIPDESQVGTTRAGVKDVILESAITTIMDFESVAAVDAADKVLGVRNWLGNKGLAA	300	
Db	242	LIPDESQVGTTRAGVKDVILESAITTIMDFESVAAVDAADKVLGVRNWLGNKGLAA	301	
Qy	301	AVDKDGFARVLNRDRNTYAPGGQFTLPGRSLMFRVNVGHLMTNDAI VDTDGSEVFEG	360	
Db	302	AVDKDGFARVLNRDRNTYAPGGQFTLPGRSLMFRVNVGHLMTNDAI VDTDGSEVFEG	361	

361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMGHPAEVFTCELSFSEVEDVLG 420
362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMGHPAEVFTCELSFSEVEDVLG 421
421 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 480
422 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 481
481 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 540
482 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 541
541 VPSPATAATHALHGHVQDVDAVAAQOGLAGKRATIEQLLTIPAKELAWAPDEIREVEDNN 600
542 VPSPATAATHALHGHVQDVDAVAAQOGLAGKRATIEQLLTIPAKELAWAPDEIREVEDNN 601
601 COSILGVVVRWVDQVGCSPVDIHDVALMEDRATLRISOLLANWLHGVITTSADVRAS 660
602 COSILGVVVRWVDQVGCSPVDIHDVALMEDRATLRISOLLANWLHGVITTSADVRAS 661
661 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 720
662 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 721
721 KARAAEKPAAPSDRAGDDAAR 740
722 KARAAEKPAAPSDRAGDDAAR 741

RESULT 2
US-09-396-347F-106
; Sequence 106, Application US/09396347F
; Patent No. 6506384
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Bellis, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: 32004-169276
; CURRENT APPLICATION NUMBER: US/09/396,347F
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 09/001,984
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-396-347F-106

Query Match 100.0%; Score 3810; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDELQ 60
2 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDELQ 61
61 AOIDKWHRRVTEPIDMDAYRQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 120
62 AOIDKWHRRVTEPIDMDAYRQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 121
121 LNARFALNAANARWGSYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP 180
122 LNARFALNAANARWGSYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP 181
181 LSSGSFGDGTFTVDQGLVVALPKSTGLANPQFAGYTCABSPSVLLINHLHIEI 240
182 LSSGSFGDGTFTVDQGLVVALPKSTGLANPQFAGYTCABSPSVLLINHLHIEI 241
241 LIDPESQVGTTRDAGVKDVLIESAITTIMDFEDSVAAADKVLGYRNWMLGNKGLDAA 300

242 LIDPESQVGTTRDAGVKDVLIESAITTIMDFEDSVAAADKVLGYRNWMLGNKGLDAA 301
301 AVDXDGTAFLEVLNRDRNNTAPGGGQFTLPGRSIMFVRNVTGHLMTNDAIYDITDGSSEVFE 360
302 AVDXDGTAFLEVLNRDRNNTAPGGGQFTLPGRSIMFVRNVTGHLMTNDAIYDITDGSSEVFE 361
361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMGHPAEVFTCELSFSEVEDVLG 420
362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMGHPAEVFTCELSFSEVEDVLG 421
421 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 480
422 LPQNTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMVRKG 481
481 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 540
482 TMSQPMILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADWVETKIAQPRAGASTAW 541
541 VPSPATAATHALHGHVQDVDAVAAQOGLAGKRATIEQLLTIPAKELAWAPDEIREVEDNN 600
542 VPSPATAATHALHGHVQDVDAVAAQOGLAGKRATIEQLLTIPAKELAWAPDEIREVEDNN 601
601 COSILGVVVRWVDQVGCSPVDIHDVALMEDRATLRISOLLANWLHGVITTSADVRAS 660
602 COSILGVVVRWVDQVGCSPVDIHDVALMEDRATLRISOLLANWLHGVITTSADVRAS 661
661 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 720
662 LERMAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGYTEPILHRRRREF 721
721 KARAAEKPAAPSDRAGDDAAR 740
722 KARAAEKPAAPSDRAGDDAAR 741

RESULT 3
US-09-252-991A-22524
; Sequence 22524, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22524
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22524

Query Match 67.5%; Score 2570; DB 4; Length 731;
Best Local Similarity 69.1%; Pred. No. 2.1e-240;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDELQ 60
8 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDELQ 67
61 AOIDKWHRRVTEPIDMDAYRQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 120
68 AOIDKWHRRVTEPIDMDAYRQFTEIGYLLPEPDDFTITTSVGDAEITTTAGPQVVPV 127
121 LNARFALNAANARWGSYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP 180
128 MNARFALNAANARWGSYDALYGTVDVIPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP 187

QY 181 LSSGSGDGTFTVQDQOLVWALPKS-TGLANPGQFAGYTGAAESPTSVLLINHLHIE 239
Db 188 LESGSHVDATSYKNGALVWALPKNGSETGLKNGAQFLAQDGAAPQAVLKLKHLHFE 247
QY 240 ILIDPESQVGTDRAGVKVILESAITTIMDFEDSVAADAAKVLGYRNWGLKNGDLA 299
Db 248 IQIDPSSPVQTDAGVKVLEAAITTIMDCEDSVAADADDKVIYRNWGLKNGDLA 307
QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNGLHMTNDAIVTDGSEVPE 359
Db 308 EGVSKGSGTFTTMRPNDRVYTRADGSELTHGRSLLFVRNGLHMTNDAIILDKDNEVPE 367
QY 360 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSIIVIVKPKMHGPAEVAFTCELSRYED 417
Db 368 GTDGLFTSLIAH-----DLNGTSRKNSRTGSIIVIVKPKMHGPEEAFTNELFRVED 422
QY 418 VGLPQNTWKIGIMDEERTTNVNLKACIKAAADRVVFINTEFLDRTGDSIHISMEAGPMV 477
Db 423 VLGLPNTLKVGTIMDEERTTNVNLKACIKAAADRVVFINTEFLDRTGDSIHISMEAGAVV 482
QY 478 RGTWKSQPKILAYEDHNVDAAGLAAFGSRAQVKGKMTMTLMADMVETKIAQPRAGAS 537
Db 483 RKGAMKSEKIGAYENNNVVDVGLATGLQKQAQIGKGMWAMPDLMAANLEQKIGHPLAGAN 542
QY 538 TAWPSPPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLTIPLAKELAMAPDEIRREV 597
Db 543 TAWPSPPTAATLHALHYHVDVFAQAELAKTRPASVDDILTIPLAPNTNWTAEIKNEV 602
QY 598 DNNQSIILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISQALLANLWLRHGVITSADV 657
Db 603 DNNQSIILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISQALLANLWLRHGVISQEV 662
QY 658 RASLERMAPLVRQNDAGVAYRPMAPNPDSDSIAFLAAQELIISGAQOQNGYTEPILHRRR 717
Db 663 VESLKEMAVVRQNASDSYRPMAPNPDNDVAFQAALELVVEGTQPNGYTEPILHRRR 722
QY 718 REFKA 723
Db 723 REFKA 728

RESULT 4

US-08-836-943-2

/ Sequence 2, Application US/08836943
/ Patent No. 5965391
/ GENERAL INFORMATION:
/ APPLICANT: Reinscheid, Dieter
/ APPLICANT: Bickmanns, Bernhard
/ APPLICANT: Sahm, Hermann
/ TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN
/ BACTERIA
/ NUMBER OF SEQUENCES: 3
/ CORRESPONDENCE ADDRESS:
/ ADDRESSES: The Firm of Karl F. Ross, PC
/ STREET: 5676 Riverdale Ave.
/ CITY: Bronx
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10471
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/836.943
/ FILING DATE: 08-MAY-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Myers, Jonathan
/ REGISTRATION NUMBER: 26,963
/ REFERENCE/DOCKET NUMBER: 20357
/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (718) 884-6600
/ TELEFAX: 718/601-1099
/ TELEX: 620428
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 739 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-836-943-2

Query Match 60.1%; Score 2289; DB 2; Length 739;

Best Local Similarity 61.3%; Pred. No. 4.1e-213;

Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

QY 1 TDRVSNGNLRIRLVLYDFVNEALPGTDIDPDSFWAGVKVADLTQONALLNARDELQ 60
Db 20 TERVDAGMQVAKVLYDFVTEAVLPRVGVDAERFWSGFAAIARDLTTRNELLARDELQ 79
QY 61 AQIDKWHRRVIEPI DMDAYRQFLTEIGYLLPBPDDTITTSVDAEITTAGPQLVVPV 120
Db 80 MLIDYHRRN-SGTIDQAYEDELKEIGYLVVEPEAAEIRTONVDTEISSTAGPQLVVPV 138
QY 121 LNARFALNANARAGSLYDALYGTDPVETDGAEGKPTYNKRGDKVIAVARFLDSDVP 180
Db 139 LNARFALNANARAGSLYDALYGTNAIPETDGAEGKPEYNPVRGQKVIENGRFLDSDVP 198
QY 181 LSSGSGDGTFTVQDQOLVWALPKS-TGLANPGQFAGYTGAAESPTSVLLINHLHIE 240
Db 199 LDGASHADVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIE 258
QY 241 LIDPESQVGTDRAGVKVILESAITTIMDFEDSVAADAAKVLGYRNWGLKNGDLA 300
Db 259 QIDVHPHIGKADTKGLKDIIVLESALTITIMDFEDSVAADAAEDKTLGYSNFGLNTGELKE 318
QY 301 AVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNGLHMTNDAIVTDGSEVPE 360
Db 319 EMSKNGRIFTRELKORVYIGRNGTELVLHGRSLLFVRNGLHMTNDAIVTDGSEVPE 377
QY 361 IMDALFTGLIAHGLKASDVNGPLIINSRTGSIIVIVKPKMHGPAEVAFTCELSRYEDV 420
Db 378 IMDAVLTTCVCAIPQAFQ--KMRNSRKGSIIVIVKPKHGPPEVAFTNELFORVEDL 434
QY 421 LPQNTWKIGIMDEERTTNVNLKACIKAAADRVVFINTEFLDRTGDSIHISMEAGPMVRKG 480
Db 435 LPRHTLKVGVWDBERTSVNLDASIMEVADRLAFINTGFLDRTGDSIHISMEAGPMVRKA 494
QY 481 TMKSQPMILAYEDHNVDAAGLAAFGSRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 540
Db 495 DMQTAPWKQAYENNVDAIGTQRLPGKAQIGKGNWAMTELMAEMLEKKIQOPREGANTAW 554
QY 541 VPSPTAATLHALHYHVDVAAVQOGL-AGKERRATIEQLTIPLAKELAMAPDEIRREV 599
Db 555 VPSPTGATLHATHYLVDFVKVQDELRAAGRRSLRNILTIPTNTNWESEBKEMDN 614
QY 600 NCOSILGYVVRWVDQVGCSCVPDIHDVVALMEDRATLRISQALLANLWLRHGVITSADV 659
Db 615 NCOSILGYVVRWVEHGVGCSCVPDIHDIDLMEDRATLRISQALLANLWLRHGVITSADV 674
QY 660 SLERMAPLVRQNDAGVAYRPMAPNPDSDSIAFLAAQELIISGAQOQNGYTEPILHRRR 719
Db 675 SLERMAPLVRQNDAGVAYRPMAPNPDSDSIAFLAAQELIISGAQOQNGYTEPILHRRR 734
QY 720 FKAR 723
Db 735 FKAK 738

RESULT 5

US-09-328-352-7603
/ Sequence 7603, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7603
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-7603

Query Match	58.4%	Score 2225;	DB 4;	Length 729;
Best Local Similarity	59.8%	Pred. No. 6.6e-207;		
Matches	432;	Conservative	185;	Indels 8; Gaps 4;

Qy	1	TDVSVGNLRIRVLYDFVNEALPGTIDPDSFWAGVKVADLPQNALNARDELQ	60
Db	10	TARIQKGLAIAXELYFIENEALPGSLDSEYWKNFQVVDLSPKAKALLAKRDELQ	69
Qy	61	AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDPTITTSQVDAEITTTAGPQLVVPV	120
Db	70	AKIDEMHNNKFE---LGAYKAFLEIGYLLPEVDFQITENVDIEALLAGPQLVVPV	126
Qy	121	LNARFALNAANARWGSYLDALYGTDTVIPETDGAEGKPTYNKVRGDKVIAYARKFLDSDVP	180
Db	127	RNARYCLNAANARWGSYLDALYGFVISEBGAEGKGYNPVRGAKVIEFAKNFLNEIFF	186
Qy	181	LSSGSGFGATGFTVQDQQLVLPD-KSTGLANPGQFAGYTGAEPSVLLINHLGHE	239
Db	187	LAQSHADATKATLEQNKLVVTKDGTGTLGAHEAQVGFNGBEANPSEVLLSNGLHV	246
Qy	240	ILIDPEQGVGTTDRAGVKDVLISAITTIMDFEDSVAADAKVILGYRNWLGKNGDLA	299
Db	247	IEIDANSPIGQDLAGVKDITLEAAVTIIQDLEDSVAADAEKVEGYRNWLGKNGTLQ	306
Qy	300	AAVDKDGTAFLRNLNRDNYTAPGGGFTLPGRSLMFVRNVGHLMTNDALVDTGGSEVPE	359
Db	307	ESIEKNGKTIIVRALNKKDREIKNLIGGTTKLHGRSLMLLRNVGHLMTNPAIL-VDGEEIFE	365
Qy	360	GIMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMGGAEVAFTCELSRVEDVL	419
Db	366	GIMDALFTPLLSIADIRSENNK---NSRKGSXYIVKPKMGGEEVAFVAFELPERAEQAL	422
Qy	420	GLPQNTKIGIMDEERTTNLAKACIAAADRVVINTGFLDRGTDBEIHSTMEAGPMVRK	479
Db	423	GLPAKSLKIGIMDEERTSVNLKNCIAAAKDRITFINTGFMDRGTDBEIHSTMEAAPVVRK	482
Qy	480	GTMSKQPIWILAYEDHNDVAGLAAGFSRAQVKGKMTWMTLMADMVETKIAQPRAGASTA	539
Db	483	EAVTKQKIAAYENRNVAIGLCKGLQKQKIGKGMWPKPDSMKDMLATKAAHPNAGASCA	542
Qy	540	WVPSPTAATLHALHYHQVDAVAAVQOGLAGKERATIEQLTIPLAKELAWAPDEIREVDN	599
Db	543	WVPSPTGAVLHAMHYHQVNVKARQDLKAEMLSLDLDLTPFPATDINWSAEBINNELEN	602
Qy	600	NCQILGVVWVNDVQGVCCSKVPIDHVALMEDATIRISSOLLANLWLRHGVITSADVRA	659
Db	603	NCQILGVVWVNDVQGVCCSKVPIDNNVGLMEDSATIRISSQHVANLWLRHGVITRQVEE	662
Qy	660	SLEMAPLVDRONAGDVAYRPMAPNFDDSIATFLAAQELILSGAQQNGYTEPIHRRRRE	719
Db	663	VLEKNAKIVDRONANDPLYKPMANFETNIAFAASDLIFKGCQPSGYTEPLHAARLK	722
Qy	720	FK 721	
Db	723	LK 724	

RESULT 6
 US-09-328-352-5503
 ; Sequence 5503, Application US/09328352

; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 5503
 ; LENGTH: 3892
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-5503

Query Match	3.6%	Score 139;	DB 4;	Length 3892;
Best Local Similarity	21.6%	Pred. No. 0.0043;		
Matches	167;	Conservative	78;	Mismatches 300; Indels 228; Gaps 41;

Qy	7	GNL-----RIARVLYDFVNEALPGTIDPDSFWAGV---DKVVADLPQNALNARDE	58
Db	1368	GNLVDGDTVTATATDPAGNTSLPGTGTVSADITAPVVALDVLNDSTP---ALGTGVND	1424
Qy	59	LOAQIDKWHRRVIEPIDMDAYRQFLTEIG-----YLLPEPDFTIT-----TS	102
Db	1425	PTA-----TVVNVVDGTDYPAVNGDGTWTLADNTLPVLADGPHITVTATDAAGNA	1476
Qy	103	GYDAEITTTAGPQLVVPVNLNARFALNAAN-----ARWGSYLDALY-GT-DVIPETD	151
Db	1477	GTDTAVVTIDITAPN--APVLD--PINATDPVSGTAEAGSTVTVTPDGTATWAGTD	1531
Qy	152	GAEGKFT-YNKVRGDKVIAYARKFL-DDSVLSSGSGFGATGFTVQDQQLVVALPDKSTG	209
Db	1532	GSWSVPENPGNLVDGDTVTATATDPAGNTSLPGTGTVSADITA-----PVVALDVLN	1584
Qy	210	LANPQCFAGYTGAEPSVLLIN-HGLH-----IETLID-PES-QVG	249
Db	1585	DSTP---ALGTGVNDPTATVVNVVDGTDYPAVNGDGTWTLADNTLPVLADGPHITVT	1640
Qy	250	TTDRAGVKDVLISAITTIMDFEDSVA-----AVDAADKVLGYRNWLGKNGDLAAAV	302
Db	1641	ATDAAG--NAGTDTAVVTI---DITAPNAPVLPDIPNATDPVSG---TAEAGSTVTVT	1689
Qy	303	DKDGTAFRLNRDNYTAPG-----GGQFTLPGRSLMFVRNVGHLMT	345
Db	1690	YDGTATVAVAGTSGWSVPNPGNLVDGDTVTATATDPAGNTSLPGTGTVSADITAPVVA	1749
Qy	346	NDAIVDTDQSEVFEG-IMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKM----	400
Db	1750	LDDVLNDSTPALGTVNDPTATVVNVVDG-----TDYPAVNGDGTWTLADNTLPALAD	1804
Qy	401	GPAEVAFTCELSRVEDVLGLPQNTKIGIMDE-----ERRTTNVLKACIAAAD---	450
Db	1805	GHITITVT-----ATDAAGNVGNDTAVVTIDTSVPVSLDLDLNTDTPALTGAIDDP	1858
Qy	451	RVVFINTGFLDRGTDBEIHSTMEAGPMVRKGTMSKQPIWILAYEDHNDVAGLAAGFSRAQV	510
Db	1859	ATVVNVVDGID-----YPATN-----N	1875
Qy	511	GKGMVMTLMADMVETKIAQPRAGASTAWPSPPTAATLHALHYHQVDAVAAVQOGLAGKR	570
Db	1876	GGGTWLDAD---NTLPALIDGPHITVTATDPAGNTATDTA-----TLTIDTVADLIG--	1926
Qy	571	RATIEQLLTIP--LAKELAWAPDEIREVDNNCOSILGYVVRVVDQGVCCSKVPIDHVA	628
Db	1927	-----AITIPEDINGDGLNADLGTGGSFNAVALG-----PDALDGT	1965
Qy	629	LMEDRAT-LRISOLLANLWLRHGVITSA---DVRASLERMAPLVDRONAGDVA	677
Db	1966	VVNVANGTNYVTAAADLAN---GYITAAIPVTGEGFVAIHAEDVADQGNVDVA	2014

RESULT 7

US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Belach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; FILE REFERENCE: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3
Query Match 3.5%; Score 135; DB 3; Length 3816;
Best Local Similarity 21.3%; Pred. No. 0.01;
Matches 190; Conservative 112; Mismatches 297; Indels 294; Gaps 49;
QY 25 PG-----TDTPD-SFWAGVDKVVADLTTPQNALNARDLQAIQDKWRRVIEPID---- 76
Db 2968 PGRFVLADIGDSASDAPRAVA-----SAASEVAIRAGAVVYPRLARATDEGL 3017
QY 77 -----MDAYRQPLTIG-----YLLPEPDDFTTTTSGVDAEITTTAGPOLVVPVNLARF 125
Db 3018 VVADEAAGPRLDVTAGLANLALVPCD-----ASRPLGPEV-----RI 3059
QY 126 ALNANARNGSLVDALYGTVDIPEDT--GAEGPTYNKY-----RGKVIAYARKFL 175
Db 3060 AVRAAGV--NFRDVLALGMYPDEGLMGAEEAGVTEVGGVTTIAPGDRVWG----- 3110
QY 176 DDSVPLSSGSGFATGFTVQDGLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINH 235
Db 3111 -----LVTFGFPV---AVTHRRMLVRMP-----RGWSFAEASVPVAFUTAYYA 3152
QY 236 LH-----IEILIDPESQVGTDRAGKVDI-----LE 262
Db 3153 LHDLAGLRGESVLVHSAAGVGMAAVQLARHWADEVFGTASGKWDVLAAGLDBEHIG 3212
QY 263 SAITIMDFEDSVAA-----VDAADKVL--GYRNWLGKNGDL---A 299
Db 3213 SSRTT--EPEQRFATSGRGHIDVNLNLSGDFVDASARELLREGGR-FVEMGKTDIRTL 3269
QY 300 AAVDKDGTAFRLVLRNDRNVTAPGG-----GQFTLP-----GR 332
Db 3270 GVVGADGVPIRVAVD---LAEGAERIGQMLDEIMALFDAGVLELPLRANPVRRAHE 3326
QY 333 SLMFV---RNUGH-LMNTDAIVDTGSEVPFEGIMDALFTGLIAHGLKASDVNGPLNRS 388
Db 3327 ALRFVSGARHGKVLTVPAALDAEGTVLTGA--CTLGALVARHLVTEHDERVLLVSR 3384
QY 389 TGSIIYVKPMHGP-----AEVAF-TCELFSR-----VEDVLGLPQNTM-----KIGIM 431
Db 3385 SG-----VAPDLAELGALGAETVVAACDVANKALKALLEDI--PPEHPTGIVITAGVL 3438
QY 432 DE-----ERRITVNLKACIAAADRVVFFINTGFLDRTGDEIHT---SMEAGFMVRKG 480
Db 3439 DCGWSGLTPERVDTV-LXPKVDAAALTLESVIGELDLDPALFVITSSAASMLGGPQGSY 3497
QY 481 TKMSQPHILAYEDHNVDAAGAFSGRAQVGKMW-----TMTLMADMVETKIAOPRAGA 536

Db 3498 AAANQ-FLDTLARHRRRGLTS-----VSLGWGLWHEASGLTGGIADIDRDMVS--RAGI 3549
QY 537 STAWVPSPATAHL-----ALHYHOVDVAVAQOGLAGKRRATIEQLL----- 578
Db 3550 A-----PMPTDEALHDFDRATELGDFVLLPMLNEAALEDRAAD---GTLPLLSGLVVR 3602
QY 579 -----TIP--LAKELAWAPD---BIRBEVDNCCQSIILGV-----VWV 612
Db 3603 HRPSARAGTATAAPATGPEAFARLAAAPDPRRALRDLVRGHVALVLCHSGPEAIDAEQA 3662
QY 613 DQGVCSKVPIHDVVALMEDRATIRISSQLLANWLHGVITSADVRSLEMAPLVDRQN 672
Db 3663 FRDIGFDSLTAVELNRNLNAETGLRLPGTLVFDYPN-----PSALADHLELLAP----- 3712
QY 673 AGDVAYREMAFNFDSDIAFLAAQELILSGAQPNGYTEPIHLHRREREFKARAA 725
Db 3713 ----ATQTTAALP---JAELEERVEQLLSAAASPGPASAVDEETRTLIATRIA 3758
RESULT 8
US-08-237-919-2
; Sequence 2, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; APPLICANT: Cepek, Karyn L
; TITLE OF INVENTION: Methods and Compositions for
; TITLE OF INVENTION: Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,919
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-237-919-2
Query Match 3.4%; Score 128; DB 1; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.004;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;
QY 91 LPBDDFTTI-----TTSGVDAEITTTAGPOLVVPVNLNARFALNANARWGSYDAL 141
Db 307 LPDKMFTINRNTGVISVITGLDRESFPTV--TLVV-----QAADIQSGGL--ST 353
QY 142 YGTDVIVPETDGAEGPTYNKY--RGD-----KVIAVARKFLDSDVPLSSGSGFCDATGPT 193
Db 354 TATAVITVTNDNPPFNFTTYKQVPENEAENVITTLKVTADAPNTAWAEVYITLN 413
QY 194 VQDGLVVAL-PDKSTGLAN-----PGQFAGYTGAAESPTSV 229

Db 414 DDGGQFVVTNPNVNDGILKTAKGLDPEAKQYILHVAVTNVPFVSLTSTATVTDV 473
QY 230 LLINHLGLHIEILIDPESQVGTDRAGV-----KDVILESAITTIMDPEDSVAAV 278
Db 474 LDVNEG---PIFVPPKREVESEDFGVGQEITSYTAQEPDTFMEQKITVRI-WRDT----- 525
QY 279 DAADKVLGYRWLGLN--KGDAAAADKDGTAFLRVLNDRNYTAPGGQFTLPGRSLMF 336
Db 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHMTNDAIVDTD--GSEVPEGIMDALFTGLIAIHGLKASDV--NGPLINRTGSIY 393
Db 560 -----LIIATDNGSPVATG-----TGTL-----LILSDVNDNAPIPEPTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTMKIGIM 431
Db 600 ERNPKQVINIHDADLPNTSPFTAELTHGRVNPNTIYNDPTQESIIILKPKMALEV-- 657
QY 432 DEERTTVNLKACIAAADRVRVFINTEGFLDRGTDEIHTSMEAGPMVRKGTMKQSPWILAY 491
Db 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVKGKMTMTLMADVMETKIAQPRAGASTAWPSPATA 547
Db 697 -----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPEPLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVQOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYVYDEGGEGBEDQDFLSQLHRGLDARPEVTRNDVAPTILMSVRYLPRPANDPEI 803
QY 594 REEVNNCQS 603
Db 804 GNFDENLKA 813

RESULT 9

US-08-732-429-2
; Sequence 2, Application US/08732429
; Patent No. 630080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,429
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-429-2

Query Match 3.4%; Score 128; DB 4; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.004;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEDDFTI-----TTSQVDAEIIITAGPOLVVPVFNARFALNAARWGSGLDAL 141
Db 307 LPDKMFTINRTGTVISVVTGDLDRSPPTV--TLVV-----QAADIQSGSL--ST 353
QY 142 YGTVIPIETDGAEGKPTNKV--RGD-----KVIAYARKFLDDSVLPSSGSGFDATGFT 193
Db 354 TATAVITVTDTNDNPPFIENFTYKQVPEANEANVVITLKVTDADAPNTPAWEAVYITLN 413
QY 194 VQDGLQVVAL-PDKSTGLAN-----PCQFAGYTGAASPTSV 229
Db 414 DDGGQFVVTNPNVNDGILKTAKGLDPEAKQYILHVAVTNVPFVSLTSTATVTDV 473
QY 230 LLINHLGLHIEILIDPESQVGTDRAGV-----KDVILESAITTIMDPEDSVAAV 278
Db 474 LDVNEG---PIFVPPKREVESEDFGVGQEITSYTAQEPDTFMEQKITVRI-WRDT----- 525
QY 279 DAADKVLGYRWLGLN--KGDAAAADKDGTAFLRVLNDRNYTAPGGQFTLPGRSLMF 336
Db 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHMTNDAIVDTD--GSEVPEGIMDALFTGLIAIHGLKASDV--NGPLINRTGSIY 393
Db 560 -----LIIATDNGSPVATG-----TGTL-----LILSDVNDNAPIPEPTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTMKIGIM 431
Db 600 ERNPKQVINIHDADLPNTSPFTAELTHGRVNPNTIYNDPTQESIIILKPKMALEV-- 657
QY 432 DEERTTVNLKACIAAADRVRVFINTEGFLDRGTDEIHTSMEAGPMVRKGTMKQSPWILAY 491
Db 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVKGKMTMTLMADVMETKIAQPRAGASTAWPSPATA 547
Db 697 -----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPEPLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVQOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYVYDEGGEGBEDQDFLSQLHRGLDARPEVTRNDVAPTILMSVRYLPRPANDPEI 803
QY 594 REEVNNCQS 603
Db 804 GNFDENLKA 813

RESULT 10

US-09-798-267-2
; Sequence 2, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Cepek, Karyn
; APPLICANT: Brenner, Michael
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; INTERACTIONS WITH T LYMPHOCYTES
; FILE REFERENCE: L0560/7008RP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-798-267-2

Query Match
Best Local Similarity 3.4%; Score 128; DB 4; Length 878;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TTSGVDABEITTTAGPQLVVPVFNARFALNAANARWGLSYDAL 141
DB 307 LPDKXMTINRNTGVISVTTGLDRESFPTY--TLVW-----QAADLQEGGL--ST 353
QY 142 YGTDVIPETDCAEKGFTYKVV--RGD-----KVIAVARKFLDDSVPLSSGSGFGDATGFT 193
DB 354 TATAVITVTDNDNPPIFNPTTKGQVPENANVITTLKVTDAAPNTPAWEAVYITLN 413
QY 194 VQDQQLVVAL-PDKSTGLAN-----PGQFAGYTGAAESPTSV 229
DB 414 DGGQFVVTTNPVNDGILKTAKGLDPEAKQVILHVAVTNVVPEVSITTTSTAIVTVDV 473
QY 230 LLINHGHLHIELIDPESQVGTTRAGV-----KDVILESAITTIMDPEDSVAAV 278
DB 474 LDVNEG---PIFVPEKREVESEDFGVQEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNMLGLN--KGDAAAADKDGTAFLVLRNDRNYTAPGGQFTLPGRSLMF 336
DB 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHLMTNDIAVDTD--GSEVPEGIMDALFTGLIAIHGLKASDV--NGPLINSTRGSIY 393
DB 560 -----LIATDNGSPVATG-----TGTL--LILSDVNDNAPIPEPRTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTWKIGIM 431
DB 600 ERNPKQVINIHADLPNTSPPTAELTHGRVNPNTIQNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKGTMSQSPWILAY 491
DB 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVGKGMWMTMELMADWVETKIAQPRAGASTAWVPSPTAA 547
DB 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAAVQOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
DB 744 TRDNVYYDEEGGEEEDQDPLSQLHRGLDARPEVTRNDVAPTILMSVPRYLPRPANPDEI 803
QY 594 REEVNNQCS 603
DB 804 GNFDENLKA 813

US-09-798-267-3

Query Match
Best Local Similarity 3.4%; Score 128; DB 4; Length 878;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY 91 LPEPDDFTI-----TTSGVDABEITTTAGPQLVVPVFNARFALNAANARWGLSYDAL 141
DB 307 LPDKXMTINRNTGVISVTTGLDRESFPTY--TLVW-----QAADLQEGGL--ST 353
QY 142 YGTDVIPETDCAEKGFTYKVV--RGD-----KVIAVARKFLDDSVPLSSGSGFGDATGFT 193
DB 354 TATAVITVTDNDNPPIFNPTTKGQVPENANVITTLKVTDAAPNTPAWEAVYITLN 413
QY 194 VQDQQLVVAL-PDKSTGLAN-----PGQFAGYTGAAESPTSV 229
DB 414 DGGQFVVTTNPVNDGILKTAKGLDPEAKQVILHVAVTNVVPEVSITTTSTAIVTVDV 473
QY 230 LLINHGHLHIELIDPESQVGTTRAGV-----KDVILESAITTIMDPEDSVAAV 278
DB 474 LDVNEG---PIFVPEKREVESEDFGVQEITSYTAQEPDTFMEQKITRYI--WRDT---- 525
QY 279 DAADKVLGYRNMLGLN--KGDAAAADKDGTAFLVLRNDRNYTAPGGQFTLPGRSLMF 336
DB 526 -----RNWLEINPDTGAISTRAELDREDFEHVKN--STYTA----- 559
QY 337 VRNVGHLMTNDIAVDTD--GSEVPEGIMDALFTGLIAIHGLKASDV--NGPLINSTRGSIY 393
DB 560 -----LIATDNGSPVATG-----TGTL--LILSDVNDNAPIPEPRTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPQNTWKIGIM 431
DB 600 ERNPKQVINIHADLPNTSPPTAELTHGRVNPNTIQNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKGTMSQSPWILAY 491
DB 658 ----DYKINLKLMDNQNKDQVTTLEVSVCDEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSGRAQVGKGMWMTMELMADWVETKIAQPRAGASTAWVPSPTAA 547
DB 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAAVQOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
DB 744 TRDNVYYDEEGGEEEDQDPLSQLHRGLDARPEVTRNDVAPTILMSVPRYLPRPANPDEI 803
QY 594 REEVNNQCS 603
DB 804 GNFDENLKA 813

RESULT 11
US-09-798-267-3
; Sequence 3, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Cepek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008ER
; CURRENT APPLICATION NUMBER: US/09798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429

RESULT 12
PCT-US95-05518-2
; Sequence 2, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; FILE REFERENCE: E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/05518
 FILING DATE: herewith
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/237,919
 FILING DATE: 3 May 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Plumer, Elizabeth R.
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B0601/7023
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 878 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-05518-2

Query Match 3.4%; Score 128; DB 5; Length 878;
 Best Local Similarity 22.0%; Pred. No. 0.004;
 Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;

QY	91	LPEPDDFTI-----TTSGVDAEITTTAGPOLVVPVLNARFALNANARWGSYDAL	141
DB	307	LPDKMFTINRTGVSIVTTGLDRESFPTY--TLVV-----QAADLQEGGL--ST	353
QY	142	YGTDVIPETDGAEGPYNKV--RGD-----KVIAARFELDSVPLSSGSGDAGTFT	193
DB	354	TATAVITVTNDNPPIFNPTTYRGQVPENEAIVITLKVTDADAENTPAWEAVYTLN	413
QY	194	VQDQQLVVAL--PKDKSTGLAN-----PGQFAGYTGAAESPTSV	229
DB	414	DDGQFVVTNPVANDGILKTAKGLDFEAKQYILHVAVTNVVPFVSLTSTATVTDV	473
QY	230	LLINHLHIELIDPESQVGTTRAGV-----KDVILESAITTIMDFEDSVAAV	278
DB	474	LDVNEG--PIFVPEKRVESDFGVGOEITSYTAQEPDTFMEQKITRYI--WRDT--	525
QY	279	DAADKVLGYRNWGLN--KGPLAAAVDKGTAFRLVLRNDRNNTAPGGGQFTLPGRLMF	336
DB	526	-----RWLEINPDTGAISTRAELDREDFEHVN--STYA-----	559
QY	337	VRNVGHMTNDAIVDTP--GSEVFGINDALFTGLIAHGLKASDV--NGPLINRTGSIY	393
DB	560	-----LIIATDNGSPVATG-----TGTL--LILSDVNDNAIPEPRITFFC	599
QY	394	IVKPK-----MHG-----PAEVAFTCEL--FSRV-----EDVLGLPQNTMKIGIM	431
DB	600	ERNPKQVINIHDADLPNTSPFTFAELTHGRVNPWTIQYNDPTQESLILKPKMALEVG--	657
QY	432	DEERTTVNLKACIKAAADRNVFINTGLDRTGDEIHTSMEAGFMVRKGMKSPWILAY	491
DB	658	-----DYKINLKLMDNQNDQVTTLEVSVCDEG-----AAGVCR-----KAQP-----	696
QY	492	EDHNVADAGL-----AAGFSGRAQVCKGNWMTLMADWVETKIAQPRAGASTANVPSTAA	547
DB	697	-----VEAGLQIPAILGILG-----GILAL--LILILLLLLFURRAVVKPELPLP-PEDD	743
QY	548	TLHALHYH-----QVDVAAVQOGLAK-----PRATIEQLLTIPLAKELAWAPDEI	593
DB	744	TRDNVYYDEBGGGEEDQDLSQLHGLDARPEVTRNDVAPTLMSVPRPLPAPNDEI	803
QY	594	REEDVNNQCS	603
DB	804	GNFIDENLKA	813

RESULT 13
 US-09-105-537-6
 ; Sequence 6, Application US/09105537A
 ; Patent No. 6265202
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600.438US1
 ; CURRENT APPLICATION NUMBER: US/09/105,537A
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 11877
 ; TYPE: PPT
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-105-537-6

Query Match 3.4%; Score 128; DB 3; Length 11877;
 Best Local Similarity 20.7%; Pred. No. 0.34;
 Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY	9	LRIARVLDFVNEALPGTDIDPDSFWAGVVKVADLTPOQALLNAADELOAQIDKWHR	68
DB	10193	LRDAGVL-----DTVLRLTGIEPEFGSGSGDGAADPGAEPEASI---DLDDEA--LI	10241
QY	69	RVIEPID-----MDAYRQELTEIGYLLPPDPDDFTITTSVDAEITTTAGPOLVVP	119
DB	10242	RNALGPRMTWSSNEQLVDALRASLKENEELKES-----RRADRRQEPMAIVG	10291
QY	120	VLNARFALNANARWGSYDAL--YGTDV---IPETDGAEGKTYNKVRGDKVIAYARK--	173
DB	10292	--MSCRFAGGIRSP--DLMDAVAAGKDLVSEVPSERGWDIDSLYDPVGRGTTTVRNA	10348
QY	174	FDDDSVPLSSGSGF---DATGFTVQDQVLVALPKSTGLANFGQFAGYTGAAESPTSV	229
DB	10349	FDDDAAGDDAAFFGLSPREALAMPDQQLLEA-----SWEVFERAGIDFASV	10396
QY	230	LLINHLHI-----EILIDPESQVGTTRAGVKDKVILESAITTIMDFEDSVAAVDA	280
DB	10397	RGTDVGVVVGCGYQDYAPDIRVAPEGTGGYV--VTGNSSAVASGRIAYSLGLEGPAVTVD	10455
QY	281	A---DKVLGYRNWGLNKGLAAAVDKGTAFRLVLRNDRNNTAPGGGQFTLPGESLMFV	337
DB	10456	ACSSSLVALHIALKGLRNGDCSTAL--VGGVAVL-----ATPGAFIEFS	10497
QY	338	-----RVNGHMTNDAIVDIDGSEV--FEGIMDALFTG--LIAIHGLKASDVNGPL	384
DB	10498	SQQAAVADGRTKGPASAADGLAWGEVAVLLERLSDAARRKGRHVAV-----VRGA	10550
QY	385	INSRTGSIYVVKPKKHGPAEVAFTCELFSEVEDVLGLPQNTMKIGIMDEERTTVNLKAC	444
DB	10551	INQDGASNLTPA--HGFSQ-----OHLIROLALD--ARLT-----	10582
QY	445	IKAAADRNVFINTGLDRTGDEIHT-----SMEAGPMVRKGMKSPWILAY	485
DB	10583	---SSDSDVVEGHGTGLDGFIEAQALLATYGGQFAGQPLRLGLTKSNIGHTQAASGV	10639
QY	486	-----PWILAYED--HNVD--AGLA-----	508
DB	10640	AGVIRKMWQALRHGVLPKTLTHVDEPTDQVDWSAGSVELLTEAVDWPFERGLRRAGVSAGF	10699
QY	509	QVKGKMTWMTLMADWVETKIAQPRAGASTA--W--VPSPTAATL-----HALHYHGV	557
DB	10700	VGTTNAHVLEAPAVESPAVEPPAGGVVVPWVPSAKTSAAALDAQIQLAAAYAEEDTDV	10759
QY	558	DVAAVQOGL-----AGRRATIEQLTIP--LAKELAWAPDEI-----	593

Db 10760 DPAVARALVDSRTAMEHRAVAVGDSREALRDLRMPEGLVRGTVDPRGAVFPGQGT 10819
QY 594 -----REEVDNN-----CQSILGYVVRVWDOGV--GCSKVPDIHDVAMEDRATL 636
Db 10820 QWAGMGNELDSSPEFAAAMAECEALSPYVDHSLAVVROAPSATILDRVDVQV--VTF 10878
QY 637 RISSQLLANWLRHGVITSADVRASLERMAPLVDRQNDAGVAVRPMAPNFDD 688
Db 10879 ANVSLAKVWQHGGITPEAVIGHSGGEIA-----AAYVAGALTLLDDA 10920
RESULT 14
US-09-252-991A-21798
; Sequence 21798, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21798
Query Match 3.3%; Score 127; DB 4; Length 1650;
Best Local Similarity 20.2%; Pred. No. 0.015;
Matches 177; Conservative 110; Mismatches 349; Indels 242; Gaps 37;
QY 7 GNLRIRALVDFVNNALPCTDIDPDSFWAGV-----DKVVADLTPOQALLNARDEL 59
Db 382 GNLAALAVVEDVQPLDQRAASTDQATLVQVAAPOVEDDALAD--OPALALHQAIDR 438
QY 60 QAQIDKWHRRRRIEIDMAYROFLTEIGYLPEDPDTITTSQVDAE-ITTTAGPQLVY 118
Db 439 QAQVAFGGDFAAVGGVDAPC-RQFQFAVA---CQQAATVVQVGGAAQORLLATEGAATV 494
QY 119 PVLNARFALNAARWGLYDALYGTDPITDQ-NEKFTYKVRGDKVIA-----169
Db 495 EV-----GAADVQ-HALADQATLGVQVORASGDAQG-----VAGEQALAAVQGGAG 541
QY 170 -----YARKFLD---DSVPLSSGSPG---DATGFTVQ-----DGQLVVAL 203
Db 542 ERQAGAGNGAFGVVQRAVDAQVDAAGAGGALGAVVQASGRDVQAGGADHPGQAVV--599
QY 204 PDKSTGLANPGQFAGYTGAAESTSVLLINH-----LH-----237
Db 600 -----QGLGDAQGAG--GADQAPAAVVQAGGEGEGAGVDFPAGAVVHRAELAQOQAG 653
QY 238 -----IEILIDPESQVGTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWGL 293
Db 654 RGDQAAVAVDQRAAIEVEGDAGFADODASAL--VEAFQVGEQALGADPSLLAVVQPG 711
QY 294 NKGD-----LAAAVDKDGTAFRLVNRDRNYTA-----321
Db 712 HOGDAGVAADAATAVAVVQAGADIHRTLGDHAGTAVWEAGALQRHAGIAEQAPALVQOR 771
QY 322 --PGGQFTLPGRSLMFVRNVGHMTNDALVTDGSEVEGIMDALFTGLIAIHGL-----375
Db 772 GLAQSQRTGAGGPRATVQVQARGQAFAFQRAALVVQCHAAHAQAQVLAVEPTAVAV 831
QY 376 -KASDVNGLPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLG-----420
Db 832 EQFAAVQAQAVAPGQHPGLGLVQALHGEAQAAVADDLAAVQLLAGVHGLRGAGNLAG 891

QY 421 ----LPQ-----NTMKIGIMDEERRITTVNLKA-----CIKAAADRVVFIINTFLDR 462
Db 892 AVVDLPRLDSDAALRGDQPLAVVDRVGRDLQGLFADQFATLLGQAKRRLQVALGGTTP 951
QY 463 TG--DEIHTSMEAGPMWKGTKMSQPMILAYEDHNVDAGLAA-----GFSGRAGVG 511
Db 952 SGVADGVRCEYCAATAEQLATLV-----VAQAGDLHAHASLAAGAAVAVWEGSAGRGAQAG 1007
QY 512 KG-MWTMT'ELMADVVETKIAQPRAGASTAWVPSP'ATAATHALHYHQVDVAAVQOGLAGKR 570
Db 1008 VGHQHAAIVVQAGAVDVQLAAARQQAAGAVVQLPHRHARLALAEQVAFIAVQQ-----1061
QY 571 RATIEQLLTIPLA-KELAWAFDEIREEVNDCQSIILGYVVRVWDOGVGCSKVPDIHDVAL 629
Db 1062 -AAEADLQAPAAPAORALAAAVVQAARPV-----EAVGCGEQAE-----AV 1100
QY 630 MEDRATLRISSQLLANWLRHGVITSADVRASLERMAPLVDRQNDAGVAVRPMAPNFDD-DS 688
Db 1101 VD--GTAGDHOQVADQLAAAIIVERTD--AGLDAL-----AGDFAVPVDORLGDADA 1148
QY 689 TAFLLAAQELIISGAQPNQNGYTEPILHRRRREFKARAAE 726
Db 1149 GGLLAAEQALLA-----VVERRADLQVVQAD 1175
RESULT 15
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117859
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1998-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1998-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-2
Query Match 3.3%; Score 126.5; DB 3; Length 3739;
Best Local Similarity 21.4%; Pred. No. 0.066;
Matches 166; Conservative 87; Mismatches 297; Indels 227; Gaps 45;
QY 24 LPGTDIDPDSFW--AGVDKVVADLTPOQALLNARDELQAOIDKWHRRRIEIPIDMAY 80
Db 1551 LPGGVAGFEDLWRLVAGEDAISGF-PQDRG-----WDVEGLYDP-DPDAS 1594
QY 81 RQFITEIGYLLPEPDDFTTITTSQVDAEITTTAGPQLVVPVNLNARFALNAARWGLSYDA 140
Db 1595 GRTYCRAGGFLDEAGEFDADFFGISPREALAMDQ-----QRLLETS---WEAVEDA 1644
QY 141 LYGVTDVIPETDGAEGKPTYNKVRGDKVIAYA-----RKFLDSDVPLSSGSGFGDTCFT 193

Db 1645 --GID-----PT--SLOGQOVGVFAGTNGPHYEPLLRNRTAEDLEGYVGTGNAAS 1689
QY 194 VQDGLVVALPDKSTGLANPCQFAGYTGAAESPTSVLLINHLHIEILIDPESQVGTDDR 253
Db 1690 IMSGRVSTL-----GLEGPV-----TVDTCSSSLV--ALHLVQALRKCECGLALA 1736
QY 254 AGVKDVLILESAITIMDP-----ED--SVAAYDAADKVLGYRNWGLNKGDLAAAVDK 304
Db 1737 GG--VTWMSPTTFVEFSRQGLAEDGRSKAFASADG-FGPAEGVGMILLVERLSDARR 1792
QY 305 DGTAFRLV-----LNRD--RNYTAPGGQFTLQRSIMFVRNVHLMNTDAIVDTDGSE 356
Db 1793 NGRVLAUVVRGSANQDASNGLTAPNG-----PSQORVIRRALADARLTADVD----- 1842
QY 357 VEGIMDALFTG-----LIAIHGLKASDVNGPL-----INSR-----TGS 391
Db 1843 ---VVEAHGTGTRLGDPLEAQUALIATVG-QGRDTEQPLRLGSLKSNIGHTQAAAAGVSGI 1897
QY 392 IYIVKPKXHGPAEVAFTCELSERVEDVLGLPQNTMKI--GIMDEERTTVNLKACIKAAA 449
Db 1898 IKMVOAMEHGVLPKTLHVD---RPSDQIDWSAGTVELLTEAMDWPRKOEGLR----- 1947
QY 450 DRUVFINTGFLDRTGDEIHTSMEAGPMVRKGTMSQ-----PMILAYEDHNVDAAGLAAG 503
Db 1948 -RAAVSFTGI---SGTNAHIVLEAPVDEADAPADEPSVGGVVPWL-----VSAKTPAA 1996
QY 504 PSGRAQVKGKMWMTTELMAVMETKIAQPRAGASTAWVPSPTAATLH---ALHYHQVDVA 560
Db 1997 LD--AQIGR-----LAAPASQGRDAADP--GAVARVLAGGQAQFEHRAVALGTGQDDLA 2047
QY 561 ---AVQOGL-----AGKERATI-----EOLLTIPLAKELAWAPDEIREVDN 599
Db 2048 AALAAPEGLVRGVASGVGRVAFVFPGGQTQWAGMGAEILLV--SKEFAAAVAE----- 2098
QY 600 NCQSTILGYVVRW-----VDQGVGCSKVPDIHDVALMEDRATLRISSELLANWLRHGVITS 654
Db 2099 -CEAALAPYVDSLEAVVRQAFG---APTLEKVDVQOP-VTFVAVVSLAKVQHHGVTPQ 2153
QY 655 ADVRASLERMAPLVDRQVAGDVAYRPMAPNFDSDI-----AFLAAQELILS 700
Db 2154 AVVGHSGGEIA-----AAYVAGALSUDDAARVVTLRSKISGAHLAGQGGNLS 2200

Search completed: November 21, 2003, 16:11:36
Job time : 18.6054 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 / Search time 44.8027 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSVGNLRIRARVLYDFVN.....KARAAEKPAFSDRAGDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	741	AAE05665	Mycobacterium tube
2	3810	100.0	1172	AAU01900	M. tuberculosis Tb
3	3804	99.8	2502	AAE00117	M. tuberculosis po
4	2289	60.1	739	AAE077501	Malate synthase.
5	2289	60.1	739	AAE03234	C glutamicum prote
6	2275	59.7	739	AAE79539	Corynebacterium gl
7	2253	59.1	712	AAE79540	Corynebacterium gl
8	991.5	26.0	1190	ABG18357	Novel human diagno
9	741.5	19.5	830	ABG25453	Novel human diagno

10	594.5	15.6	504	22	ABG18358	Novel human diagno
11	585.5	15.4	1641	22	ABG18356	Novel human diagno
12	433.5	11.4	441	22	ABG18355	Novel human diagno
13	420	11.0	130	23	ABP03007	Human ORFX protein
14	189.5	5.0	533	13	AAE37585	E.coli malate synt
15	189.5	5.0	533	24	AAO16702	Escherichia coli a
16	135	3.5	3816	21	AAE92708	S. antibioticus 8,
17	134.5	3.5	1440	23	ABBS4801	Lactococcus lactis
18	130	3.4	1624	23	ABU05888	M. tuberculosis an
19	128.5	3.4	823	22	ABG05169	Novel human diagno
20	128.5	3.4	823	22	ABG26009	Novel human diagno
21	128	3.4	878	15	AAE55060	Sequence of human
22	128	3.4	878	16	AAE85487	Human E-cadherin p
23	128	3.4	878	23	ABB1475	Human E-cadherin p
24	128	3.4	12199	21	AAE77180	S. venezuelae pik
25	127.5	3.3	5701	22	ABG51536	Human liver peptid
26	127.5	3.3	5701	22	ABG36684	Peptide #4190 enco
27	127.5	3.3	5701	22	ABG22021	Protein #4020 enco
28	127.5	3.3	5701	22	AAE57448	Human brain expres
29	127.5	3.3	5701	22	AAE59843	Human bone marrow
30	127.5	3.3	5701	22	AAE05328	Peptide #4010 enco
31	127.5	3.3	26926	22	AAU05396	Human titin (conne
32	127.5	3.3	31267	24	ABG74786	Human RAS11 protei
33	126.5	3.3	3739	21	AAE18638	Amino acid sequenc
34	126.5	3.3	3739	21	AAE77193	S. venezuelae macr
35	126.5	3.3	3739	21	AAE77201	S. venezuelae pik
36	126.5	3.3	3739	21	AAE77202	Narbolonide syntha
37	126.5	3.3	3739	24	ABG71662	S. venezuelae narb
38	125.5	3.3	2039	19	AAE56322	Haemophilus paraga
39	125	3.3	918	20	AAE98422	Chlamydia pneumoni
40	124	3.3	1346	21	AAE77195	S. venezuelae macr
41	124	3.3	4999	23	AAO22158	Ramoplanin biosynt
42	124	3.3	15281	15	AAE44929	T. niveum Cyclospo
43	123	3.2	1346	21	AAE18640	Amino acid sequenc
44	123	3.2	1346	21	AAE77204	Narbolonide syntha
45	123	3.2	1346	24	ABG71664	S. venezuelae narb

ALIGNMENTS

RESULT 1

AAE05665

ID AAE05665 standard; peptide; 741 AA.

XX AC AAE05665;

XX DT 24-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis 88 kDa secreted protein.

XX KW Mycobacterial disease; infection; 88-kDa protein; tuberculosis; TB;

XX LW early antigen; human immunodeficiency virus; HIV; secreted protein.

XX OS Mycobacterium tuberculosis.

XX PN US6245331-B1.

XX PD 12-JUN-2001.

XX PF 31-DEC-1997; 97US-0001984.

XX PR 02-JAN-1997; 97US-0034003.

XX PA (UNIV-) UNIV NEW YORK MEDICAL CENT.

XX PA (COLS) UNIV COLORADO STATE.

XX PI Laal S, Zolla-Pazner S, Belisle JT;

XX DR WPI; 2001-424324/45.

XX PT Detecting a mycobacterial disease (tuberculosis) in individuals

PT comprise assaying a biological sample for the presence of

PT anti-Mycobacterium tuberculosis antibodies or M. tuberculosis
 XX antigen-antibody complex
 PS Claim 14; Column 121; 96pp; English.
 XX
 CC The present invention relates to a method for early detection of active
 CC mycobacterial disease or infection comprising assaying a biological fluid
 CC sample for the presence of early antibodies specific for an 88-kDa
 CC Mycobacterium tuberculosis (Mtb) protein or immune complexes consisting of
 CC an 88-kDa M. tuberculosis protein antigen complexed with an antibody
 CC specific for the antigen. The method is useful for the early and rapid
 CC detection of mycobacterial disease, particularly tuberculosis, in
 CC individuals at heightened risk of developing tuberculosis. This
 CC individuals include human immunodeficiency virus (HIV)-infected subjects
 CC or other immunocompromised individuals. The method is a rapid and
 CC inexpensive screening procedure for detecting mycobacterial disease.
 CC The present sequence is a Mtb 88 kDa secreted protein early antigen,
 CC used in the exemplification of the invention.
 XX
 SQ Sequence 741 AA;
 Query Match 100.0%; Score 3810; DB 22; Length 741;
 Best Local Similarity 100.0%; Pred. No. 4,7e-316;
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSTWAGVGVKVVADLTPOQALLNARDELQ 60
 DB 2 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSTWAGVGVKVVADLTPOQALLNARDELQ 61
 QY 61 AQIDKWHRRVIEPIDMDAYEQFTEIGYLLPEPDDFTITTSQVDAITTAGQLVVPV 120
 DB 62 AQIDKWHRRVIEPIDMDAYEQFTEIGYLLPEPDDFTITTSQVDAITTAGQLVVPV 121
 QY 121 LNARFALNAARWGSGLYDALYGVTDVTPETDGAEGKPTYNKVRGDKVIARFELDDSV 180
 DB 122 LNARFALNAARWGSGLYDALYGVTDVTPETDGAEGKPTYNKVRGDKVIARFELDDSV 181
 QY 181 LSSGSFGDGTFTVQDGLVVALPKSTGLANPQFAGYTGAAESPTSVLLNHLHIEI 240
 DB 182 LSSGSFGDGTFTVQDGLVVALPKSTGLANPQFAGYTGAAESPTSVLLNHLHIEI 241
 QY 241 LIDPESQVGTTRAGVXDVILESAITTIMDEDSVAADKVLGYRNWGLNKGDLAA 300
 DB 242 LIDPESQVGTTRAGVXDVILESAITTIMDEDSVAADKVLGYRNWGLNKGDLAA 301
 QY 301 AVDKDGTAFRLVNRDRNYTAPGGGQFTLPGRSLMFRVNVGHLMTNDAIVDTGSEVFEG 360
 DB 302 AVDKDGTAFRLVNRDRNYTAPGGGQFTLPGRSLMFRVNVGHLMTNDAIVDTGSEVFEG 361
 QY 361 IMDALFTGLIAIHGLKASDVNGPLINSGTGIYIVKPMHGPFAEVAFTCELSFVEDVLG 420
 DB 362 IMDALFTGLIAIHGLKASDVNGPLINSGTGIYIVKPMHGPFAEVAFTCELSFVEDVLG 421
 QY 421 LPONTMKIGIMDEERTTVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 480
 DB 422 LPONTMKIGIMDEERTTVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 481
 QY 481 TMKSQPTWILAYEDHNVDAGLAGSGRAQVKGKWWMTLMADVETKIAQPRAGASTAW 540
 DB 482 TMKSQPTWILAYEDHNVDAGLAGSGRAQVKGKWWMTLMADVETKIAQPRAGASTAW 541
 QY 541 VPSPTAATHALHYHVDVAVVQOGLAGKRATTEQLLTIPLAKELAWAPDEIRREVDNN 600
 DB 542 VPSPTAATHALHYHVDVAVVQOGLAGKRATTEQLLTIPLAKELAWAPDEIRREVDNN 601
 QY 601 QOSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQALLANWLHGHVITSADYRAS 660
 DB 602 QOSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQALLANWLHGHVITSADYRAS 661
 QY 661 LERMAPLVDQRNAGVAVRPAWPNFDDSIAPLAAQELTSLGAQOPNGVTEPILHRRREF 720
 DB 662 LERMAPLVDQRNAGVAVRPAWPNFDDSIAPLAAQELTSLGAQOPNGVTEPILHRRREF 721

QY 721 KARAAEKPAFSDRAGDDAAR 740
 DB 722 KARAAEKPAFSDRAGDDAAR 741
 RESULT 2
 AAU01900
 ID AAU01900 standard; Protein; 1172 AA.
 XX AAU01900;
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis Tbf14 fusion protein.
 XX Tbf14; antigen; vaccine; tuberculosis; AIDS; His tag; Mtb81; Mo2;
 KW acquired immunodeficiency disease.
 XX Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 3..8
 FT /label= Histidine_tag
 FT /note= "Nickel chelating region used for purifying
 FT the fusion protein"
 FT Region 9..749
 FT /label= Mtb81_region
 FT Misc-difference 49
 FT /note= "Encoded by TC"
 FT Region 750..1172
 FT /label= Mo2_region
 XX WO200124820-A1.
 PD 12-APR-2001.
 XX
 PF 10-OCT-2000; 2000WO-US28095.
 XX
 PR 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 PA (CORI-) CORIXA CORP.
 FI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Iodes ML;
 XX WPI; 2001-290576/30.
 DR N-PSDB; AA303794.
 XX
 PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 PS Claim 3; Fig 3; 168pp; English.
 XX
 CC The sequence represents Mycobacterium fusion protein antigen Tbf14
 CC consisting of a His tag for purification, antigen Mtb81 and antigen Mo2.
 CC Compositions comprising at least 2 heterologous antigens as a fusion
 CC protein, and vectors expressing the fusion proteins are used as vaccines
 CC to prophylactically immunise mammals (especially humans) against
 CC infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 SQ Sequence 1172 AA;
 Query Match 100.0%; Score 3810; DB 22; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 9.3e-316;
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSTWAGVGVKVVADLTPOQALLNARDELQ 60
 DB 9 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSTWAGVGVKVVADLTPOQALLNARDELQ 68

FT Misc-difference 1035 /note= "Encoded by TAA stop codon"

FT Misc-difference 1082 /note= "Encoded by TGA stop codon"

FT Misc-difference 1160 /note= "Encoded by TAA stop codon"

FT Misc-difference 1161 /note= "Encoded by TGA stop codon"

FT Misc-difference 1164 /note= "Encoded by TAA stop codon"

FT Misc-difference 1187 /note= "Encoded by TAA stop codon"

FT Misc-difference 1188 /note= "Encoded by TAA stop codon"

FT Misc-difference 1211 /note= "Encoded by TGA stop codon"

FT Misc-difference 1253 /note= "Encoded by TGA stop codon"

FT Misc-difference 1344 /note= "Encoded by TAA stop codon"

FT Misc-difference 1348 /note= "Encoded by TAA stop codon"

FT Misc-difference 1386 /note= "Encoded by TAA stop codon"

FT Misc-difference 1527 /note= "Encoded by TAA stop codon"

FT Misc-difference 1579 /note= "Encoded by TGA stop codon"

FT Misc-difference 1582 /note= "Encoded by TGA stop codon"

FT Misc-difference 1616 /note= "Encoded by TGA stop codon"

FT Misc-difference 1628 /note= "Encoded by TGA stop codon"

FT Misc-difference 1643 /note= "Encoded by TGA stop codon"

FT Misc-difference 1673 /note= "Encoded by TAA stop codon"

FT Misc-difference 1683 /note= "Encoded by TAA stop codon"

FT Misc-difference 1685 /note= "Encoded by TAA stop codon"

FT Misc-difference 1691..2438 /note= "Encoded by TAA stop codon"

FT Protein /label= Mtb-81 polypeptide

FT Misc-difference 2439 /note= "Encoded by TAG stop codon"

FT Misc-difference 2440 /note= "Encoded by TGA stop codon"

FT Misc-difference 2464 /note= "Encoded by TAA stop codon"

FT Misc-difference 2471 /note= "Encoded by TGA stop codon"

FT Misc-difference 2478 /note= "Encoded by TGA stop codon"

FT Misc-difference 2489 /note= "Encoded by TAA stop codon"

XX WO200055194-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07196.

XX 18-MAR-1999; 99US-0272975.

XX (CORI-) CORIXA CORP.

XX Hendrickson RC, Lodes MJ, Houghton RL;

XX WPI; 2000-638180/61.

XX N-PSDB; AAA53971.

XX Novel Mycobacterium tuberculosis polypeptide comprising an immunogenic

PT

portion of M. tuberculosis antigens Mtb-81 and Mtb-67.2, useful for diagnosis, treatment and monitoring therapy of tuberculosis

Claim 1: Fig 1a-f; 9lpp; English.

Polypeptides comprising an immunogenic portion of Mycobacterium tuberculosis antigens Mtb-81 or Mtb-67.2 or variants that differ by substitutions, additions, insertions and/or deletions but which still react with antigen specific antisera or T-cells are described. Also described are polynucleotides encoding the polypeptides. The polypeptides, expression vectors expressing the polypeptides or comprising an antisense polynucleotide, or an antigen presenting cell comprising a sequence encoding the polypeptides are useful for determining the presence or absence of M. tuberculosis in whole blood, serum, sputum, plasma, saliva, cerebrospinal fluid or urine in a patient infected with human immunodeficiency virus (HIV).

SQ Sequence 2502 AA;

Query Match 99.8%; Score 3804; DB 21; Length 2502;

Best Local Similarity 99.7%; Pred. No. 9.3e-315;

Matches 738; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDRSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVGVVADLTPOQALLNARDLQ 60

Db 1699 TDRSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVGVVADLTPOQALLNARDLQ 1758

QY 61 AOIDKWHRRVTEPIDMDAYRQFLTEIGVLLPEPDDFTITTSVDAAETTTAGPOLVVPV 120

Db 1759 AOIDKWHRRVTEPIDMDAYRQFLTEIGVLLPEPDDFTITTSVDAAETTTAGPOLVVPV 1818

QY 121 LNARFALNAANARWGLYDALYGTDTVITPDGAEKGYTNKVRGDKVIAYARKFLDDSV 180

Db 1819 LNARFALNAANARWGLYDALYGTDTVITPDGAEKGYTNKVRGDKVIAYARKFLDDSV 1878

QY 181 LSSGSPGDATGTTVDGQGLVVALPDKSTGLANPGFAGYTGAAESPTSVLLINHLHIEI 240

Db 1879 LSSGSPGDATGTTVDGQGLVVALPDKSTGLANPGFAGYTGAAESPTSVLLINHLHIEI 1938

QY 241 LIDPESQVTTDRAGVKDVILESATITIMDFDSVAADKVLGYRNWGLNKGDLAA 300

Db 1939 LIDPESQVTTDRAGVKDVILESATITIMDFDSVAADKVLGYRNWGLNKGDLAA 1998

QY 301 AYDKGTAFLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNDALVDTGSEVFEG 360

Db 1999 AYDKGTAFLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNDALVDTGSEVFEG 2058

QY 361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFACTELPSRVEDVLG 420

Db 2059 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFACTELPSRVEDVLG 2118

QY 421 LPQNTWKIGIMDEERTTVNLKACIKAAADRVVFTNGTFLDRTGDEIHTSMEAGPMWRKG 480

Db 2119 LPQNTWKIGIMDEERTTVNLKACIKAAADRVVFTNGTFLDRTGDEIHTSMEAGPMWRKG 2178

QY 481 TKMSQPMILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADVMETKIAQPRAGASTAW 540

Db 2179 TKMSQPMILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADVMETKIAQPRAGASTAW 2238

QY 541 VPSPTAATLHALHYHQVDVAQVQGLAGKRATITBOLLTIPLAKELANPADSIREVDNN 600

Db 2239 VPSPTAATLHALHYHQVDVAQVQGLAGKRATITBOLLTIPLAKELANPADSIREVDNN 2298

QY 601 CQSILGVVVRWVDQGVGCKVPDIHDVALMEDRATILRISSQLLANWLRHGVITSADVRAS 660

Db 2299 CQSILGVVVRWVDQGVGCKVPDIHDVALMEDRATILRISSQLLANWLRHGVITSADVRAS 2358

QY 661 LERMAPLVDQRNAGDVAYRPMAPNFDDSTAFIAAQLILSGACQPNGYTEPILHRRREF 720

Db 2359 LERMAPLVDQRNAGDVAYRPMAPNFDDSTAFIAAQLILSGACQPNGYTEPILHRRREF 2418

QY 721 KAPAAEKPAKPSDPAAGDAAAR 740

Db	2419 KARAEEKPAPSDRAGDDAAR 2438
XX	AAAR77501
XX	AAAR77501 standard; Protein; 739 AA.
XX	AAAR77501;
XX	04-JUN-1996 (first entry)
XX	Malate synthase.
XX	aceB gene; malate synthase; Coryneform bacterium; regulatory region; protein synthesis.
XX	Coryneform glutamicum.
XX	DB4440118-C1.
XX	09-NOV-1995.
XX	11-NOV-1994; 94DE-4440118.
XX	11-NOV-1994; 94DE-4440118.
XX	(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX	Eikmanns B, Reinscheid D, Sahn H;
XX	WPI; 1995-374802/49.
XX	N-PSDB; AAT05501.
XX	DNA contg. upstream regulatory region from malate synthase gene of coryneform bacteria - used for regulated synthesis of protein in coryneform(s), regardless of nutrient medium compsn..
XX	Example 3; Page 7-8; 12pp; German.
XX	AAAR77501 is encoded by an approx. 3 kb BfrI-PvuII fragment contg. the aceB gene from Coryneform glutamicum (see AAT05501) Nucleotides 1 to 574 can be isolated and used to regulate the expression of a protein-encoding structural gene placed downstream of it, after incorporation into a vector and transfer into a Coryneform bacterium. Induction of protein expression occurs even when the culture medium contains carbon sources other than acetate.
XX	Sequence 739 AA;
XX	Query Match 60.1%; Score 2289; DB 16; Length 739;
XX	Best Local Similarity 61.3%; Pred. No. 3.1e-186;
XX	Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;
XX	1 TDRVSVGNLRIRARLYDFVNEALPGTDIDPDPSFWAGVDKVVADLTQONALLNARDELQ 60
XX	20 TERVDAGNOVAKVLYDFVTEAVLPRVGVDKFKWSGFAAARDLTPRNRELLARRDELQ 79
XX	61 AQIDKWHRRVIEPIDMDAYRFLTEIGVLPPEPDDFTITTSVGVDAEITTTAGPQLVVPV 120
XX	80 MLTIDYHRRNN-SGTIDQEAIVEDFLXIGVLEPEAEIRTONVDTEISSTAGPQLVVPV 138
XX	121 LNARFALNANARWGLSDYALYGTVDVITPDGAEGKPTYNKRGDKVIYARKELDSVVP 180
XX	139 LNARFALNANARWGLSDYALYGTWAIPTPDGAEGKEYNPRVGOKVIEWGREFLDSVVP 198
XX	181 LSSGSFGDGTFTVDGQLVVALPDKSTGLANPGOFAGYTGAAESPTSVLLINGLHIEI 240
XX	199 LQASHADVKEYNITDGKLAAGHIGSVYLVKXNRESYRGFTGNFLDPEALLETNGLHIEL 258
XX	241 LIDPESQVTTBRAGVKDVLBSATTTIMDFEDSVAADVADKVLGYRNWGLNKGDLAA 300
XX	259 QIDPVHPIGKADTKGLKIVLESATTTIMDFEDSVAADVADKVLGYRNWFLNGLTGLKE 318

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium, and identifying a homologue of a gene derived from corynebacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
CC European Patent Office.

[illegible]

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbonyl, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX SQ Sequence 712 AA;
 Query Match 59.1%; Score 2253; DB 22; Length 712;
 Best Local Similarity 61.2%; Pred. No. 3.5e-183;
 Matches 438; Conservative 89; Mismatches 183; Indels 6; Gaps 4;
 QY 9 LRIARVLYDFVNNALPGTDIDPDSFWAGVKKVADLTQONALLNARDLQAIQDKWHR 68
 DB 1 MQVAKVLYDFVTEAVLPRVGVDAEKFSWGAARADLTTPNRRELLAERDELQMLIDYHR 60
 QY 69 RRVIEPIDMDAYRFLTEIGVLLPEPDDFTITTSVDAEITTTAGPOLVVPVFNARFALN 128
 DB 61 NN-SGTTDQEAEDFLKEIGVLEPEEAERTONVTEISSTAGPOLVVPVFNARFALN 119
 QY 129 AANARWGLSLYDALYGTDIVPETDCAEKGPTVKNVRGDKVAYARKFLDDSVPLSSSGFGD 188
 DB 120 AANARWGLSLYDALYGTDIVPETDCAEKGPTVKNVRGDKVAYARKFLDDSVPLSSSGFGD 179
 QY 189 ATGTVDQGVQVLPALPKSTGLANPGQFAGYTGAAESPTSVLLINHLHLEIILIDPESQV 248
 DB 180 VEKYNITDGLAAHIGSVYLRKRESYRGFTGNFLDPEAILLETNGLHLEIILIDPESQV 239
 QY 249 GTTRPAGVKDVLIESAITTMDPDSVAAYDAADKVLGYRNWGLNKGDLAAAVDGDGA 308
 DB 240 GKADKTLGDKLVLESAITTMDPDSVAAYDAADKVLGYRNWGLNKGDLAAAVDGDGA 299
 QY 309 FLRVNLRDNTYAPGGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSSEVFEGIMDALFTG 368
 DB 300 FTRELKDRVIGRNGTELVHGRSLLFVRNVGHLMQNPSIL-IDGEEIFEGIMDAVLT 358
 QY 369 LIAIHGLKASVNGPLNSRGTGSIYIKPKHGAPEAVAFCELPFSRVEDVLGLPQNTMKI 428
 DB 359 VCAIPGIAFQV---KMRNSRKGSIYIKPKHGAPEAVAFCELPFSRVEDVLGLPQNTMKI 415
 QY 429 GIMDEERTTNLXACIKAAADRVVFTNGTFLDRTGDEIHTSMAGMVRKGTMSQPMI 488
 DB 416 GIMDEERTTNLXACIKAAADRVVFTNGTFLDRTGDEIHTSMAGMVRKGTMSQPMI 475
 QY 489 LAYEDHNVDAAGLAFSGRAQVKGWMTMELMDVETKIAQPRAGASTAWPSPAT 548
 DB 476 QAYENNNVDAGTQRLPGKAQIGKGMWMTMELMDVETKIAQPRAGASTAWPSPAT 535
 QY 549 LHALHYHVDVAAVQOGL-ACKRRATIEQLITIFLAKELAWAPDEIRREEDNNQCOSILGY 607
 DB 536 LHALHYHVDVAAVQOGL-ACKRRATIEQLITIFLAKELAWAPDEIRREEDNNQCOSILGY 595
 QY 608 VYRVDOGVGSKVPDIHVALMEDRATLRISSQLANWLRHGVITISADVRASLERVAPL 667
 DB 596 VYRVDOGVGSKVPDIHVALMEDRATLRISSQLANWLRHGVITISADVRASLERVAPL 655
 QY 668 VDRQAGVAYRPMAPNFDSDIFLAQELTILSGAQQPQNGVTEPILHRRRREPKAR 723
 DB 656 VDRQAGVAYRPMAPNFDSDIFLAQELTILSGAQQPQNGVTEPILHRRRREPKAR 711

RESULT 8
 ABG18357
 ID ABG18357 standard; Protein; 1190 AA.
 XX AC ABG18357;
 DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #18348.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX PA Drmanac RT, Liu C, Tang YT;
 XX PI WPI: 2001-639362/73.
 XX DR N-PSDB; AAS82544.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX PS Claim 20; SEQ ID No 48716; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1190 AA;
 Query Match 26.0%; Score 991.5; DB 22; Length 1190;
 Best Local Similarity 53.1%; Pred. No. 3.6e-75;
 Matches 195; Conservative 66; Mismatches 97; Indels 9; Gaps 3;
 QY 56 RDELOAQIDKWHRRVIEPIDMDAYRQFLTEIGVLLPEPDDFTITTSVDAEITTTAGPQ 115
 DB 187 RDRIOQAALDEWHRRNPGPVKDKAAYKSFRELGLVLPQPERVTVETTGIDSEITSQAGPQ 246
 QY 116 LVVPLNARFALNANARWGLSLYDALYGTDIVPETDGAEGKGTYNKVRGDKVAYARFAL 175

Db 247 LVVPAMNARYALNAANRWGSLYDALYSDIIPQEGAMVSG--YDPQGEQVIAWRRFL 304
QY 176 DSVFLPSGSGFDGATGFTVQDQVLVALPD-KSTGLANPGFAGVTGAESPTSLLINH 234
Db 305 DESLPLENGSYODVAFKVDKQLRIQLKNGKEITLRTPAQFVGVRGDAAPTILVKN 364
QY 235 GLHIEILDPESQVQTTDRAGVKOVILBSAITTIMDFDSVAADAAVKVLGYRNWGLN 294
Db 365 GLHIGVQIDANGRIYKQDPFGHINDGIVETAISTILDCEDSVAADAEKILLYRNLLGLM 424
QY 295 XGDIAAAVDKGTAFRLVLRNDRNTAPGGGQFTLPGRSLMFVRNVGHMTNDALVDTG 354
Db 435 QOTLQEKKEKGRQIVRKLNDRHTAADGSEISLHGRSLFIRNVGHMTIPVWDSG 494
QY 355 SEVFEGINDALFTGTHIAHGLKASDVNGPLINSRTGSIYIVKPKKHGPAEVAFTCELSR 414
Db 485 NEIPEGILDGVMGTGAIYDLKVQK-----NSRTGSYIVKPKKHGQOEVAFAFANKLFR 538
QY 415 VEDVLGL 421
Db 539 IETMLGI 545

RESULT 9
ABG25453
ID ABG25453 standard; Protein; 830 AA.
XX AC ABG25453;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #25444.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS89640.
XX CC New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX PS Claim 20; SEQ ID NO 55812; 103pp; English.

The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 830 AA;

Query Match 19.5%; Score 741.5; DB 22; Length 830;
Best Local Similarity 60.3%; Pred. No. 4.5e-54;
Matches 146; Conservative 34; Mismatches 53; Indels 9; Gaps 3;
QY 412 FSRV--EDVLGLPQNTWKIGINDEERTTVNLKACIKAAADRVVFTNGFLDRTGDEIHT 469
Db 398 FSKMAGHDTIVLGTETVYKMGINDEERTTVNLKACIKAAADRVVFTNGFLDRTGDEIHT 457
QY 470 SMEAGPMVRKGTMKSQPWILAYEDHNVDAAGLAAFGSGRAQVKGVMWMTMELMADVETKI 529
Db 458 VMEAGPMVRKGTMKSQPWILAYEDHNVDAAGLAAFGSGRAQVKGVMWMTMELMADVETKI 517
QY 530 AQRAGASTAWVSPPTAATLHLYHVDVAAVQOGLA-----GKRRTIEQLLTPLAK 584
Db 518 DQLRAGANTAWVSPPTAATLHLYHVDVAAVQOGLA-----GKRRTIEQLLTPLAK 577
QY 585 ELAWAPEIREEVDNNQCSTILGYVWRVWQVGCCKVPDIHDVVALMEDRATLRISSQLA 644
Db 578 NANWSAQETQOELDNNVQVGLGVVWRVWQVGCCKVPDIHDVVALMEDRATLRISSQLA 635
QY 645 NW 646
Db 636 EW 637

RESULT 10
ABG18358
ID ABG18358 standard; Protein; 504 AA.
XX AC ABG18358;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #18349.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS82545.
XX CC New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX PS Claim 20; SEQ ID No 48717; 103pp; English.

XX PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 504 AA;

Query Match 15.6%; Score 594.5; DB 22; Length 504;

Best Local Similarity 50.9%; Pred. No. 7.6e-42;

Matches 115; Conservative 40; Mismatches 68; Indels 3; Gaps 2;

QY 9 LRIRARVLYDFVNNALPCTDIDPSFWAGVYKVVADLTQNCALLNARDLQADKWHR 68

DB 139 LRIDANFRFVDEVLPTGLDAAAFWRNFEIYHDLAPENRQLLAERDRIQALDSEWHR 198

QY 69 RRVIEPIDMDAYROFLTEIGYLLPEPDFTITTSVDAEITTTAGPQLVVPVNLNARFALN 128

DB 199 SNPGPVKDKAAYKSFRLGVLVFPQPERVITETGIDSEITSQAGPQLVVPVNNARYALN 258

QY 129 AANARWGLYDALYGTVDVPIETDGAEGKPTYNKVRGDKVIAYARKFLDSDVPLSSGSGFD 188

DB 259 AANARWGLYDALYGSDDIIPQEGAMVSG--YDFQGEQVIAWRRFLDESPLNGSYQD 316

QY 189 ATGFTVQDQGLVWALPD-KGTGLANPGQFAGYTGAAESPTSVLLIN 233

DB 317 VVAFKVLQ-----IDANGRI 362

RESULT 11

ABG18356

ID ABG18356 standard; Protein; 1641 AA.

XX AC ABG18356;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18347.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX XX 30-MAR-2001; 2001WO-US08631.

XX XX 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS82543.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX PS Claim 20; SEQ ID No 48715; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1641 AA;

Query Match 15.4%; Score 585.5; DB 22; Length 1641;

Best Local Similarity 42.4%; Pred. No. 2.6e-40;

Matches 122; Conservative 46; Mismatches 71; Indels 49; Gaps 3;

QY 9 LRIRARVLYDFVNNALPCTDIDPSFWAGVYKVVADLTQNCALLNARDLQADKWHR 68

DB 942 LRIDANFRFVDEVLPTGLDAAAFWRNFEIYHDLAPENRQLLAERDRIQALDSEWHR 1001

QY 69 RRVIEPIDMDAYROFLTEIGYLLPEPDFTITTSVDAEITTTAGPQLVVPVNLNARFALN 128

DB 1002 SNPGPVKDKAAYKSFRLGVLVFPQPERVITETGIDSEITSQAGPQLVVPVNNARYALN 1061

QY 129 AANARWGLYDALYGTVDVPIETDGAEGKPTYNKVRGDKVIAYARKFLDSDVPLSSGSGFD 188

DB 1062 AANARWGLYDALYGSDDIIPQEGAMVSG--YDFQGEQVIAWRRFLDESPLNGSYQD 1119

QY 189 ATGFTVQDQGLVWALPD-KGTGLANPGQFAGYTGAAESPTSVLLINHGHLHILIDPSQV 248

DB 1120 VVAFKVLQ-----IDANGRI 1133

QY 249 GTTDRAGVKDVILESAITTMDFDSVAADDAADKVLGYRNWLGKNG 296

DB 1134 GKNDPAHINDVIVEAAITILDENSGRAVNPFIKSV-VPHLLGLGK 1180

RESULT 12

ABG18355

ID ABG18355 standard; Protein; 441 AA.

XX AC ABG18355;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18346.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX N-PSDB; AAS82542.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 48714; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 441 AA;
 Query Match 11.4%; Score 433.5; DB 22; Length 441;
 Best Local Similarity 55.9%; Pred. No. 3.4e-28;
 Matches 81; Conservative 26; Mismatches 37; Indels 1; Gaps 1;
 QY 9 LRIARVLYDFVNAELPGTDIDPDSFWAGVKVADLTTPQNALNARDELQAOIKDWHR 68
 DB 283 LRIDANFKFVDEVLPGTGLDAARFNRFDIVHDLAPENQLAERDRIQALDEWHR 342
 QY 69 RRVEIPIDMDARQFTEITGYLLPEPDDTTITSGVDABITTAGPOLVVPVNLNARFALN 128
 DB 343 SNFGPVKDKAAYKSFRLGILYLPQPERVTVTTGIDTSITSGAGPOLVVPANARYALN 402
 QY 129 AANARWGSYDALYGTVDVIPETDGA 153
 DB 403 AANARWGSYDALYGSIDIIPQ-EGA 426
 RESULT 13
 ABP03007
 ID ABP03007 standard; Protein; 130 AA.
 XX ABP03007;
 AC

XX 24-JUN-2002 (first entry)
 XX Human ORFX protein sequence SEQ ID NO:5996.
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.
 XX Homo sapiens.
 XX WO200192523-A2.
 XX 06-DEC-2001.
 XX 29-MAY-2001; 2001WO-US10836.
 XX 30-MAY-2000; 2000US-206132P.
 XX 29-AUG-2000; 2000US-228716P.
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach MD;
 XX WPI: 2002-106308/14.
 XX N-PSDB; ABN18759.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX Disclosure; SEQ ID 5996; 1037pp; English.
 XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 130 AA;
 Query Match 11.0%; Score 420; DB 23; Length 130;
 Best Local Similarity 59.7%; Pred. No. 7.9e-28;
 Matches 77; Conservative 24; Mismatches 28; Indels 0; Gaps 0;
 QY 1 TDRVSVGNLRITARYLYDFVNAELPGTDIDPDSFWAGVKVADLTTPQNALNARDELQ 60
 DB 2 TERVQGSGLQVAKVLYDFVVEEAVGAGVDAETFWAGDKVITELAPNDLLAKRDQLQ 61

QY 61 AQIDKWHRRVIEBIDMAYRQFLTEIGYLLPSPDDFTITTSVGDVABITTTAGPQLWVPV 120
 Db 62 AQIDQWRDHRASLEAVAYKAFLEIGYLLPTPAEPSVNTANVDTEITTAGPQLWVPI 121

QY 121 LNARFALNA 129
 Db 122 LNARFALNA 130

RESULT 14
 AAR37585
 ID AAR37585 standard; Protein; 533 AA.
 XX AAR37585;
 AC AAR37585;
 XX 25-MAR-2003 (updated)
 DT 29-JUL-1993 (first entry)
 XX E.coli malate synthase expressed from MTAceB2 construct.
 DE Growth Hormone; metallothionein Ia; inducible promoter;
 XX bacterial gene expression; malate synthase; glyoxylate cycle;
 KW transgenic animal.
 KW Escherichia coli.
 XX WO9218635-A1.
 XX 29-OCT-1992.
 XX 13-APR-1992; 92WO-AU00164.
 XX 16-APR-1991; 91AU-0005664.
 XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 FA Brownlee AG, Nancarrow CD, Ward KA;
 XX WPI; 1992-382118/46.
 DR N-PSDB; AAQ43182.
 XX Gene expression cassette for high level expression of bacterial
 PT and/or plant genes - comprises inducible promoter and
 PT 3'-non-coding sequence of growth hormone exon 5
 XX Example 2; Fig 6; 8pp; English.
 PS Construct MTAceB2 consists of the sheep metallothionein-Ia gene
 CC promoter sequence joined to the coding sequence of the E.coli aceB
 CC gene at a unique SalI site. This sequence was then joined by the
 CC 3' sequence of exon 5 of sheep growth hormone gene. Minor sequence
 CC modification around the initiation codon of the bacterial aceB gene
 CC was made by site-directed mutagenesis. The construct was transferred
 CC to transgenic mice and the mice were analysed for expression of
 CC malate synthase. The value for expression (nmoles product formed/mg
 CC protein/20 min) was 34.3, c.f. 0 in the control cell line.
 CC See AAQ43177-Q43183.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 533 AA;
 SQ Query Match 5.0%; Score 189.5; DB 13; Length 533;
 Best Local Similarity 22.7%; Pred. No. 3.le-07;
 Matches 120; Conservative 67; Mismatches 253; Indels 89; Gaps 21;

QY 144 TDVIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVPLSSGSGDGTGTVQDGLWVAL 203
 Db 2 TEQATTTDELAFTEPYGEQEKQLTAAVEFLTELVE-----THTPQNKLLAAR 51
 QY 204 PDKSTGLANPQFAGYTGAAEPTSVLLINHHIILIDPESQVGTTRAGVKVILES 263
 Db 52 IQQQQDIDN-GTLPDFDIFSETASIRDADWKIRGIPADLDRRVEITGVKRWKVINAN 110

QY 264 AITTIMPDEDSVAADAAADKVLGYRNMLGNKGLAAAVDKDGTAFLE-VNLRDRNYTAP 322
 Db 111 VKVFADPEDSLAP-----DW---NK-----VIDGQINLRDAVNGTISYTN 149

QY 323 GGGQFTLPGRSLMFVRNVGHLMTNDIAVDTDGSEVFEGIMDALET-GLIAIHGKA--SD 379
 Db 150 AGKIYQLKPNPAVLICVRGJHLPEKHVTWEG-----EAPGSLDFDLYFFHNYQALLAK 205

QY 380 VNGPLINSRTGSIYIVKPKMHGPAEVAFTCELFSEVEDVLGPQNTMKIGIMDEERTTV 439
 Db 206 GSGP-----YFYLPTQSQWQEAASWSEVFSVAEDRENLPRTIKATLLIETLPAVF 256

QY 440 NLKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMV---RKGTMSQPMILAYED--- 493
 Db 257 QMDEILHALRDHIIVGLNCRWDYIFSYIKTLKPNYDRLVLPDRQAVTWKDFLNAYSKLLI 316

QY 494 ---HNVDA---GLAAGPSGRAQVKGKGMWTMTLMADMVETKIAQPRAGASTAWPSPSTA 546
 Db 317 KTKHKGAFAMGMAAFIPSKDEHNQ-VLNKVKAD-----KSLEANNHGDGTWIAHGL 371

QY 547 ATLHALHYHQVDA-AVQOGLAGKRRATIEOL--LTIPLAKELAWAP---DEIREVDNN 600
 Db 372 A-----DTAMAVFNILGSRKNQLEVMREQDAPITADQLLAPCDGERTEGMRAN 421

QY 601 COSILGYVVRVVDQGVGCKVPDIHDVVALMEDRATLRISSQLLANWLRLH 649
 Db 422 IRVAVQYIEAWI-SGNGCVPI-----YGLMEDAATAEISRTSIWQNIHH 464

RESULT 15
 AAO16702
 ID AAO16702 standard; Protein; 533 AA.
 XX AAO16702;
 AC AAO16702;
 XX 29-MAY-2003 (first entry)
 DT Escherichia coli aceB protein.
 DE Enzyme; L-amino acid preparation; malate synthase A; aceB; L-threonine;
 KW human medicine; pharmaceutical industry; foodstuffs; animal nutrition.
 XX Escherichia coli.
 OS WO2003008604-A2.
 XX 30-JAN-2003.
 PD 03-JUL-2002; 2002WO-EP07352.
 XX 18-JUL-2001; 2001DE-1035051.
 PR 23-JUL-2001; 2001US-306867P.
 XX (DEGS) DEGUSSA AG.
 XX Hermann T;
 PI WPI; 2003-248014/24.
 XX N-PSDB; AAL52112.
 DR Preparing L-amino acids, e.g. L-threonine, by fermenting microorganisms
 PT of Enterobacteriaceae family in which the aceB gene is attenuated in
 PT particular eliminated, and isolating L-amino acid from culture medium -
 XX Disclosure; Page 31-33; 36pp; English.
 PS The invention comprises a method of preparing L-amino acids. The method
 XX involves fermenting microorganisms of Enterobacteriaceae family that
 CC produce the desired L-amino acid, in which the aceB (malate synthase A)
 CC gene is eliminated. The method is useful for preparing L-amino acids
 CC (e.g. L-threonine). The L-amino acids produced by the method of the
 CC invention are useful in human medicine and in the pharmaceuticals

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 28.8617 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSGNLRARVLYDFVN.....KARAEKPAISDRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	748	11 US-09-272-975-2	Sequence 2, Appli
2	3810	100.0	753	11 US-09-272-975-58	Sequence 58, Appl
3	2289	60.1	739	10 US-09-738-626-6388	Sequence 6388, Ap
4	184	4.8	530	15 US-10-158-761-9582	Sequence 9582, Ap
5	145.5	3.8	541	15 US-10-158-761-9538	Sequence 9538, Ap
6	135	3.5	3816	11 US-09-808-880-3	Sequence 3, Appli
7	132.5	3.5	5245	12 US-10-329-079-45	Sequence 45, Appl
8	130	3.4	1624	16 US-10-080-170-539	Sequence 539, App
9	128	3.4	878	9 US-09-905-983-48	Sequence 48, Appl
10	128	3.4	878	14 US-10-165-049-2	Sequence 2, Appli
11	128	3.4	878	14 US-10-165-049-3	Sequence 3, Appli
12	128	3.4	11877	10 US-09-861-289-6	Sequence 6, Appli
13	128	3.4	11877	10 US-09-860-846-6	Sequence 6, Appli
14	128	3.4	11877	11 US-09-836-821-6	Sequence 6, Appli
15	128	3.4	11877	12 US-10-271-889-49	Sequence 49, Appl

16	128	3.4	12199	11 US-09-988-384B-6	Sequence 6, Appli
17	127.5	3.3	5701	9 US-09-864-761-57319	Sequence 37319, A
18	127.5	3.3	5701	12 US-10-029-386-32438	Sequence 32438, A
19	127.5	3.3	26926	10 US-09-759-508B-2	Sequence 2, Appli
20	126.5	3.3	3739	10 US-09-861-289-33	Sequence 33, Appl
21	126.5	3.3	3739	10 US-09-860-846-33	Sequence 33, Appl
22	126.5	3.3	3739	11 US-09-988-384B-33	Sequence 33, Appl
23	126.5	3.3	3739	11 US-09-836-821-33	Sequence 33, Appl
24	126.5	3.3	3739	11 US-09-793-708-2	Sequence 2, Appli
25	126.5	3.3	3739	12 US-10-201-365-3	Sequence 3, Appli
26	126.5	3.3	3739	12 US-10-160-539-2	Sequence 2, Appli
27	126.5	3.3	3739	12 US-10-271-889-33	Sequence 33, Appl
28	125.5	3.3	2039	15 US-10-193-584-7	Sequence 7, Appli
29	124.5	3.3	3939	15 US-10-156-761-10434	Sequence 10434, A
30	124	3.3	1346	10 US-09-861-289-37	Sequence 37, Appl
31	124	3.3	1346	10 US-09-860-846-37	Sequence 37, Appl
32	124	3.3	1346	11 US-09-988-384B-37	Sequence 37, Appl
33	124	3.3	1346	11 US-09-836-821-37	Sequence 37, Appl
34	124	3.3	1346	12 US-10-271-889-37	Sequence 37, Appl
35	124	3.3	4999	10 US-09-976-059-14	Sequence 14, Appl
36	123	3.2	1346	11 US-09-793-708-4	Sequence 4, Appli
37	123	3.2	1346	12 US-10-201-365-5	Sequence 5, Appli
38	123	3.2	1346	12 US-10-160-539-4	Sequence 4, Appli
39	122	3.2	821	12 US-10-174-677-28	Sequence 28, Appl
40	122	3.2	4999	10 US-09-976-059-15	Sequence 15, Appl
41	119.5	3.1	7257	14 US-10-014-717-5	Sequence 5, Appli
42	119	3.1	642	15 US-10-156-761-9781	Sequence 9781, Ap
43	118.5	3.1	915	12 US-10-193-764-35	Sequence 35, Appl
44	118.5	3.1	1222	12 US-10-193-764-37	Sequence 37, Appl
45	118.5	3.1	1228	12 US-10-193-764-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-272-975-2
; Sequence 2, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 748
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-2

Query Match	100.0%;	Score	3810;	DB	11;	Length	748;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	740;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	TDRVSGNLRARVLYDFVNNEALPGTIDPDSFWAGVKVADLTQONQALLNARDELQ	60				
Db	9	TDRVSGNLRARVLYDFVNNEALPGTIDPDSFWAGVKVADLTQONQALLNARDELQ	68				
QY	61	AQIDKWHRRVIEPIDMDAYEQFLTEIGYLLPEPDDFTTITSGVDAEITTTAGPOLVVPV	120				
Db	69	AQIDKWHRRVIEPIDMDAYEQFLTEIGYLLPEPDDFTTITSGVDAEITTTAGPOLVVPV	128				
QY	121	LNAPALNAARVGSVLYDALYGTDIVPETDGAEGKPTYNKVRGDKVIYARKFLDSDVP	180				
Db	129	LNAPALNAARVGSVLYDALYGTDIVPETDGAEGKPTYNKVRGDKVIYARKFLDSDVP	188				
QY	181	LSSSGFGDGTGTVDQGVVVALPKSTGLANPGQFAGYTGAEPSPTSVLLINHLHLEI	240				

```
Db 189 LSSGSGDATGFTVQDGLVVALPKSTGLANPCQFAGYTGAAESPTSVLLINHLHIEI 248
Qy 241 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 300
Db 249 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 308
Qy 301 AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIVDTDSEVFEG 360
Db 309 AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIVDTDSEVFEG 368
Qy 361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELSRVEDVLG 420
Db 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELSRVEDVLG 428
Qy 421 LPQNTMKIGINDEERRITVNLKACIKAAADRVPINTGFLDRTGDEIHTSMEAGPMVRKG 480
Db 429 LPQNTMKIGINDEERRITVNLKACIKAAADRVPINTGFLDRTGDEIHTSMEAGPMVRKG 488
Qy 481 TMKSQPMILAYEDHNVDAGLAAGSFGRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 540
Db 489 TMKSQPMILAYEDHNVDAGLAAGSFGRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 548
Qy 541 VPSPTAATLHALHVQDVVAAVQOGLAGKRATIEQLTTIPLAKELAWAPDEIREEDVNN 600
Db 549 VPSPTAATLHALHVQDVVAAVQOGLAGKRATIEQLTTIPLAKELAWAPDEIREEDVNN 608
Qy 601 QOSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 660
Db 609 QOSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 668
Qy 661 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTPILHRRRREF 720
Db 669 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTPILHRRRREF 728
Qy 721 KARAAEKPAFSDRAGDDAAR 740
Db 729 KARAAEKPAFSDRAGDDAAR 748

RESULT 2
US-09-272-975-58
; Sequence 58, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 21021.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-58

Query Match 100.0%; Score 3810; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDRSVGNLRIARVLVDFVNEALPGTIDIDPDSFVAGVKVADLPQNALNARDELQ 60
Db 14 TDRSVGNLRIARVLVDFVNEALPGTIDIDPDSFVAGVKVADLPQNALNARDELQ 73
Qy 61 AQIDKWHRRVIFIDMDAYRQELTEIGYLLPPEDDFTITTSVGDVAEITTTAGPQLVWVP 120
Db 74 AQIDKWHRRVIFIDMDAYRQELTEIGYLLPPEDDFTITTSVGDVAEITTTAGPQLVWVP 133
```

```
Qy 121 LNAREALNAANRWGSLYDALYGTQDVIPTDCAEKGPTYNKVRGDKVIAYARKFLDDSV 180
Db 134 LNAREALNAANRWGSLYDALYGTQDVIPTDCAEKGPTYNKVRGDKVIAYARKFLDDSV 193
Qy 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPCQFAGYTGAAESPTSVLLINHLHIEI 240
Db 194 LSSGSGDATGFTVQDGLVVALPKSTGLANPCQFAGYTGAAESPTSVLLINHLHIEI 253
Qy 241 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 300
Db 254 LIDPESQVGTTRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWLGKNGDLAA 313
Qy 301 AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIVDTDSEVFEG 360
Db 314 AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIVDTDSEVFEG 373
Qy 361 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELSRVEDVLG 420
Db 374 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELSRVEDVLG 433
Qy 421 LPQNTMKIGINDEERRITVNLKACIKAAADRVPINTGFLDRTGDEIHTSMEAGPMVRKG 480
Db 434 LPQNTMKIGINDEERRITVNLKACIKAAADRVPINTGFLDRTGDEIHTSMEAGPMVRKG 493
Qy 481 TMKSQPMILAYEDHNVDAGLAAGSFGRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 540
Db 494 TMKSQPMILAYEDHNVDAGLAAGSFGRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 553
Qy 541 VPSPTAATLHALHVQDVVAAVQOGLAGKRATIEQLTTIPLAKELAWAPDEIREEDVNN 600
Db 554 VPSPTAATLHALHVQDVVAAVQOGLAGKRATIEQLTTIPLAKELAWAPDEIREEDVNN 613
Qy 601 QOSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 660
Db 614 QOSILGYVVRVWDQGVGCKVPDIHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 673
Qy 661 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTPILHRRRREF 720
Db 674 LERMAPLVDRONAGDVAYRPNAPNFDSDIAFLAAQELILSGAQQPNGYTPILHRRRREF 733
Qy 721 KARAAEKPAFSDRAGDDAAR 740
Db 734 KARAAEKPAFSDRAGDDAAR 753

RESULT 3
US-09-738-626-6988
; Sequence 6988, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6988
```



```
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 245-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9538
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9538

Query Match      3.8%; Score 145.5; DB 15; Length 541;
Best Local Similarity 20.1%; Pred. No. 0.00017;
Matches 101; Conservative 67; Mismatches 210; Indels 125; Gaps 21;

QY 249 GTTDRAGVKDVILSAITTIMDFDSVAANDADKVLYGNWGLNKGDLAAAVDKDGTGTA 308
Db 96 GPTDKMTINALNSGARVWLDAFEDAGAP-----TWENVVYQGL-----N 135
QY 309 FLRVLRNDRNTAGGGG-FTLPGR---SLMFVENVGLHMTNDIVDTDGSEVFEGTMDA 364
Db 136 LTDAYTRNIDFTDPKSKSYALKANEELATVVMRPRGWHLNRHLVDSQGRPVFGALVD- 194
QY 365 LFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLPQN 424
Db 195 -FGLYFFHNARR-----LLDLGKGP-YFVLPKXTESHLERLWNVFVFAQDVYGIPOG 245
QY 425 TMKIGIMDEERTTVNLKACIKAAADRVVFINTCFDRTGDEIHTSNEAGPMV-----RK 480
Db 246 TVRATVLETTTAYEEMEEILYELRDHASGLNAGRWDYLFISVKNFRDGGAKFVLPRNA 305
QY 481 TMKQPMILAYED-----HNVDAGLAGFSGRAQVGKGMWMTMTLMADMVETKI 529
Db 306 VTMTAPFMRAYTELLVTRCHKRGAAH-GGMAFIPSRDEEVKNVAFKVKAD-----KD 360
QY 530 AQPAGASTAVWPPTAATHALHYHQVDVAAVQOGLAGKRRATIEQLTIPLAKELAWA 589
Db 361 REAGDGFDSWVAHPDLV-----PIAMASFDVGLGDK-----PNQK----- 396
QY 590 PDEIREVD-----NNQCSILGVVVRWVDGVGCSKVVDIHDV 627
Db 397 -DLREDVSVAPGLLIAIDSLAETPTYNGLVNAVQVIRVIEAWL-RGLGAVAI-----F 449
QY 628 ALMEDRATLRISSQLLANWRHGVITSADVRA-----SLRMAPLVDRQAGDVAYR 679
Db 450 NLMEATAETAIERSQIQWINAGVFEFENGKATPELARKVAABELANL--RAELGBEAF- 506
QY 680 PMAPNFDDSIATLAAQELILSGA 702
Db 507 -AAGWQQ-----AHDLLLVKA 522

RESULT 6
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-2029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/429,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16

; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

Query Match      3.5%; Score 135; DB 11; Length 3816;
Best Local Similarity 21.3%; Pred. No. 0.04;
Matches 190; Conservative 112; Mismatches 297; Indels 294; Gaps 49;

QY 25 PG-----TDIDPD-SFWAGVKVVDLTPQALLNARDELQAOIDKWHRRVIEPID--- 76
Db 2968 PCRFVLADIDGDEASWDALPRAVA-----SASEVAIRAGAVVPELARATDEGL 3017
QY 77 -----MDAYRQFLTEIG-----YLLPEPDDFTITTSGVDAEITTTAGPQLWVPVLNARF 125
Db 3018 VVADEAAGPWLVDVTEAGTLANLALVPCPD-----ASRPLGPDEV-----RI 3059
QY 126 ALNANARWGSYDLYGTDIVPETD--GAEGKPTYNKV-----RGDKVIAYARKFL 175
Db 3060 AVRAAGV---NFRDYLALGMYPDEGLMGAEAAAGVTVTEVGGVTTLAPGDRVMG- 3110
QY 176 DSDVSLSSGSGFDATGFTVQDQQLVVALPDKSTGLNPGQFAGVTGAESVTSVLLNHG 235
Db 3111 -----LVTTGGFV---AVTHRMILVRNP-----RWSFAEAAASVPVAFETAYA 3152
QY 236 LH-----IBILDPSQVQTTDRAGVKQVI-----LE 262
Db 3153 LHDLAGLGGESVLVHSAAGGVGMAAVQARHWDAEVFGTASGKGDVLAQAQGLDEEHIG 3212
QY 263 SAITTIMPFDSVAA-----VDAADKVL--GVENWGLNKGDL---A 299
Db 3213 SSRIT--EFQRFRATSGRGIDVVNLNLSGDFVDASARLLREGGR-FVEMGKTDIRDDL 3269
QY 300 AAVDXDGTAFULRVLRNDRNTAPGG-----GOFTLP-----GR 332
Db 3270 GVVGADGVDPDIRYVAFD---LAEGAERIQOMLDEIMALFDAGVLRPLPRAWPVRAHE 3326
QY 333 SLMFV---RNVGH-LMTNDAIVDTGSEVFGIDALFTGLIAHGLKASDVNGPLNSR 388
Db 3327 ALRFVSOARHVGVKVVLTVPALDAEGTVLITGA--GTLGALVARHLVTEHDVRRLLVSR 3384
QY 389 TGSIVIVKPKMHGP-----AEVAF--TCELFSR-----VEDVLGLPQNTM-----KIGIM 431
Db 3385 SG-----VAPDLAAELGALGAETVVAACDVANRKAALKALLEDI--PPEHPVTGIVHTAGVL 3438
QY 432 DE-----ERRTTVNLKACIKAAADRVVFINTCFDRTGDEIHT--SWEAGPMVRKG 480
Db 3439 DDGVVSGLTPTPRVDTV-LKPKVDAALTLESVIGELDLPALFVIFSSAASMLGGFGGGSY 3497
QY 481 TMKQPMILAYEDHNVDAGLAGAAGFSGRAQVGKGMW---TWELMADMVETKIAPRAGA 536
Db 3498 AAANQ-FLDTLARHARRGLTS-----VSLQWGLWHEASGLTGGLADIDRDRMS--RAGI 3549
QY 537 STAWVSPSTAATLH-----ALHYHQVDVAAVQOGLAGKRRATIEQL----- 578
Db 3550 A-----PMPTDEALHFDRAATELGPVLLPMRLNEAALEDRAAD---GTLPELISGLVRVR 3602
QY 579 -----TIP--LAKELAWAPD---BIREVDNNCOSILGVV-----VRWV 612
Db 3603 HRPSARAGTATAAATGPEAFARELAAPDPRLRALDLVRGHVALVHSGHSGPEAIDAEQA 3662
QY 613 DQGVGCSKVPIHDVVALMEDRATIRISSQLLANWRHGVITSADVRSASLERMAPLVDRQN 672
Db 3663 FRDIGFDSLTAVELNRNRLNAETGLRLPGLTVFDYPN-----PSALADHLLLELLAP----- 3712
```

QY 673 AGDVAVRMAPNFDSDIAFLAAQELILSGAQOPNGYCTEPILHRREREFKARAA 725
Db 3713 ----ATQPTAAPL---LAELERVEQLLSAAASPGGPASAVDEBETLIATRIA 3758
RESULT 7
US-10-329-079-45
; Sequence 45, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanouel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 5245
; TYPE: PRT
; ORGANISM: Streptomyces refuineus
US-10-329-079-45
Query Match 3.5%; Score 132.5; DB 12; Length 5245;
Best Local Similarity 22.0%; Pred. No. 0.12;
Matches 179; Conservative 85; Mismatches 260; Indels 291; Gaps 47;
QY 70 RVIEPTDMDAYQFLTEIGYLLP-----PPDDFTITTSVDIAEITTAGPQLVV 118
Db 4125 RVLEAV-VEAPGTALGEIDVLLPGERELLAGASWPPDPGFVTTAGAAAD----- 4172
QY 119 PVLNARFALNANARSGSLYDALYGTDPETDGAEKGYTNKVR-GDKVIAYARKFLDD 177
Db 4173 -----GIRPPDIFEA-----QAARTHAPAVRDGGEEVAYAE----- 4204
QY 178 SVPLSSGS-----PGDATGFTVQDQLVLPKSTGLTGNPGFAGVTGA-----AB 224
Db 4205 ---LNSRANRLARLLAGRGAG---PEDTVAVLLP-RGAGLITVALVAVQKAGAAVYPLDAE 4257
QY 225 SPTSVLLINHGLHIEILIPESQVGTTRDAGVKDVLBSA--ITTIMD--FEDSVAADA 280
Db 4258 LPTG-----RIAHMLDQAKFVLTVLTGMRDALPAGAGPVVCLDDPATRAALAGLDG 4309
QY 281 ADKVLGYRNLGKGLAAAVDKGTAFRLVNLNRDNYTAPGGG---FTLPGRSL--M 335
Db 4310 ADCTDADR-----APAGDRDPAYV-----YTSGSTGTPKGVVVEQSRSLA 4352
QY 336 FVRNVCHLMTNDIAVDTDGSEVFEGIMDALFT-----GLIAHGLKASDVNCP----- 383
Db 4353 LVRSAAARYRGAAGTVLLHGSAPFDLTVTTLFPLVAGGCVIVADLDAEAGDAPNRPDLK 4412
QY 384 -----LINSRTGSIYIVKPKM-----HGPAEVAFTCE 410
Db 4413 VTPSHLAFLDGIASWAAPADLVVGGELTGLAPLRAHPCMRVYNDYGTPTATVSCA 4472
QY 411 LFRSERVEDLGPONTWKIG-----IMDEERTVNLKACIKAAADRVPFI-----N 456
Db 4473 DFV-LFEPGDELPAADVIGRPLAGHRLFVLDERLP-----VPAGVPGELYIAGVGA 4524
QY 457 TGFPLDRGTDEIHTSMEA-----GP--MVRKGTMSQPMILAYEDHNVDAGLAAGF 504
Db 4525 RGYLGRPG-----MTAERFVCGPFGPGGERMVRTGDJAR--W---RPGNLE-----Y 4567
QY 505 SGRAGVQKGMWMTTELMDMVKIA---QPRAGAS-----TAWPSPPTAATLHAL 552
Db 4568 LGR---GDGQKVKRFRIEPGEIEAALLDRFEIGQAQVAVLGRBLVAVYAAPEA----- 4618
QY 553 HYHQVDVAAVQOGLAGKR-----RATIEQLTTPLA-----KELAWAPDEIR 594
Db 4619 ---BFDPAALREGLAARLPYMWPAATVRLDALPLAPGGKLDHRLAPPEPPADAPHDRR 4675

QY 595 BEVDNQCQILGYVVRWDQVGCSS---KVPDIHDVALMEDRATL---RISSQLLANWLR 648
Db 4576 PPRD-----AW--ERVLCFAFRVGLGVAEVCAGDDFFALGSDSGICSLQVGRVR 4722
QY 649 H--GVITSADVRSLEMAP--LVDRQNAQDVAY-----RPMAPNFDSDIAFLAAQELIL 699
Db 4723 RAGGRTVDRV---FEERTPAALAARSQSGAAFEVLGGRATGVPVPTPISSWLAE---L 4776
QY 700 SGACQPNGYTEPILHRREREFKARAAEKAPSDRA 734
Db 4777 GGAAB--GYNQSVLLR-----VPAQADEA 4798
RESULT 8
US-10-080-170-539
; Sequence 539, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 539
; LENGTH: 1624
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-539
Query Match 3.4%; Score 130; DB 16; Length 1624;
Best Local Similarity 20.9%; Pred. No. 0.029;
Matches 135; Conservative 87; Mismatches 216; Indels 208; Gaps 31;
QY 153 AEKG-PTYNKVRGDKVIAYARKFLDDSVPLSSGSFG---DATGFTVQDQLVVALPKST 208
Db 916 ADKGTATFSDIANDVAKSYG-FWLGDAP-ASGSGVGYDHKAMGITAR----- 960
QY 209 GLANPGFAGVTGAESPTSVLLINHLHI-EILIDPESQ-----VGTTRAGVKDV--- 259
Db 961 -----GAWAVKR-----HFRIGIDTQDTFTVVGIDMSG---DVFNG 998
QY 260 -ILESATTIMDFE-----DSVAADAAKVLGYRNLGKGLAAAVDKGTAF 310
Db 999 MLKSHIRLIAAFDHRHIFLDPNEDAAVSWAER-----RMFELPRSSWS----- 1043
QY 311 RVLNRDNYTAPGGGQFTLPGRSLMFRNVCHLMTNDIAVDTDGSE-----VFEG 361
Db 1044 ---DYDRSLISEGGGVYSREOKAIPLSAQVRAVLGIDGSDGGAEMAPPNLIBAILRAP 1100
QY 362 MDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHPABEVAFTCELSRVEDVLGL 421
Db 1101 VDLFPNGIGIY-----IKASESDADV-----DRANDPVRV 1133
QY 422 PONTMKIGIMDEERTTVNLKACIKAAADRVPF-----INTGFLDRTG-----DEIHT 469
Db 1134 NANOVRAKVIPEGGNLGV-----TALGRVEFDLSGGRINTDALDINSAGVDCSDHEVNI 1186
QY 470 SMEAGPMVRKGTMSQPMILAYEDHNVDAGLAAGFSGRAQVKGKGMWMTTELMDMVKIA--ET 527
Db 1187 KILIDLSVSGATVKADE-----RTQJ---LESMTDEVAQLVLADN 1233
QY 528 KIAOPRAGASAWVPSPPTAATLHALHYHQVDVAAVQOGLAGK-----BRATIEQ 576
Db 1224 EDQNDLNGTSA---NAASILLPVHAWQIKYLVARGVNRELEALPSEKSIARRSEAGI 1278
QY 577 LLTTP-LAKELAWAPDEIRREVDNNQCQILGYVVRWDQVGCSSKVPDIHDVALMEDRAT 635

Db 600 ERNPKQVINIHADLPNTSPFTAEIHTHGRVNPNTIQYNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERRITVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPMVRKGTMSQPMILAY 491
Db 658 ----DYKINKLMDNQNKQDVITILEVSVCDCEG-----RAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSRAQVKGKNTWMTLMADMVETKIAQPRAGASTAWPSPATA 547
Db 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYYDEEGGEEDQDFLSQLHRGLDARPEVTRNDVAPILMSVPRYLPRANPDEI 803
QY 594 REEVDNQCOS 603
Db 804 GNFDENLKA 813
RESULT 11
US-10-165-049-3
; Sequence 3, Application US/10165049
; Publication No. US20020192724A1
; GENERAL INFORMATION:
; APPLICANT: Cepek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; TITLE OF INVENTION: Interactions with T Lymphocytes
; FILE REFERENCE: 100560/70010ERP
; CURRENT APPLICATION NUMBER: US/10/165,049
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: USSN 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: USSN 08/732,429
; PRIOR FILING DATE: 1996-11-01
; PRIOR APPLICATION NUMBER: USSN 09/798,267
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-165-049-3

Query Match 3.4%; Score 128; DB 14; Length 878;
Best Local Similarity 22.0%; Pred. No. 0.016;
Matches 134; Conservative 66; Mismatches 210; Indels 200; Gaps 33;
QY 91 LPEPDDFTI-----TTSGVDAEITTTAGPQLVVPVNLARFALNAANARWGSILYDAL 141
Db 307 LPDKNFTINRTGVISVTTGLDRESFTY--TLVV-----QAADLOGEGI--ST 353
QY 142 YGTDVIPETDGAEGKPTYNK--RGD-----KVIAAYARKFLDSDVPLSSGSGFGDATGFT 193
Db 354 TATAVITVTDNDNPPIFNPTTYKGOVPEANAVVITLKVTDADAPNTPAWEAVVTILN 413
QY 194 VQDQQLVVAL-PDKSTGLAN-----PCGAGYTGAAESPTSV 229
Db 414 DCGQFVVTNPVNNNGILKAKGLDPEAKQOYILHVATNVVPPEVSVTSTATTATVDDV 473
QY 230 LLINHLGHIILIDPESQVGTDDRAGV-----KDVILESAITTIMDFEDSVAAV 278
Db 474 LDVNEG--PIFVPPPEKREYVESDFGQGEITSYTAQEPDTFMEQKITVRI--WRJT---- 525
QY 279 DAADKVLGVNWLGLN--KGDIAAADVKGDTAFRLVNLNRDNYTAPGGQFTLPQRSIMF 336
Db 526 -----RNWLEINPDTCATRAELDRDREDFEHVN--STYTA----- 559
QY 337 VRNVGHMTNDALVDTD--GSEVFEGIMDALFTGLIAIHGLKASDV--NGPLINSRTGSIY 393

Db 560 -----LIATDNGSPVATG-----TGTL---LILSDVNDNAFIPEPRTIFFC 599
QY 394 IVKPK-----MHG-----PAEVAFTCEL-FSRV-----EDVLGLPONTMKIGIM 431
Db 600 ERNPKQVINIHADLPNTSPFTAEIHTHGRVNPNTIQYNDPTQESIIILKPKMALEVG-- 657
QY 432 DEERRITVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPMVRKGTMSQPMILAY 491
Db 658 ----DYKINKLMDNQNKQDVITILEVSVCDCEG-----AAGVCR-----KAQP----- 696
QY 492 EDHNVDAGL-----AAGFSRAQVKGKNTWMTLMADMVETKIAQPRAGASTAWPSPATA 547
Db 697 ----VEAGLQIPAILGILG-----GILAL--LILILLLLFLRRRAVVKPELLP-PEDD 743
QY 548 TLHALHYH-----QVDVAVOOGLAGK-----RRATIEQLLTIPLAKELAWAPDEI 593
Db 744 TRDNVYYDEEGGEEDQDFLSQLHRGLDARPEVTRNDVAPILMSVPRYLPRANPDEI 803
QY 594 REEVDNQCOS 603
Db 804 GNFDENLKA 813
RESULT 12
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6
Query Match 3.4%; Score 128; DB 10; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;
QY 9 LRIARVLYDFVNNEALPGTDIDPDSFWAGVDKVVADLTPQOALLNARDELQAOIDKWHR 68
Db 10193 LRDRAGVL-----DVLVLTGIEPEFGSGSDGGNADPGAEPEASI---DLDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQFLTEIGYLLPDPDDFTTTTSGVDABEITTTAGPQLVVP 119
Db 10242 RMALGPRTMTSSNEQLVDALRASLKENEELRKES-----RRRDRRQEPMAIVG 10291
QY 120 VLNARFALNAANARWGSILYDAL-YGTDV---IPETDGAEGKPTYNKYRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSPB--DLWDAAAGKOLVSEVPBERGWDIDSLYDPFGKGTIVVNA 10348
QY 174 FLDSVPLSSGSGF-----DATGFTVQDQQLVVALPDKSTGLANPGQFAGYTGAAESPTSV 229
Db 10349 FLDDAAGFADAAFFGISPREALAMPDQROLLEA-----SWEVERAGIDPASV 10396
QY 230 LLINHLGHI-----EILIDPESQVGTDDRAGVDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTGVVYVCGCYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGSIAYSLGLEGPATVDT 10455
QY 281 A---DKVLGVNWLGLNKGDLAAADVKGDTAFRLVNLNRDNYTAPGGQFTLPQRSIMFV 337
Db 10456 ACSSSLVALHLKGLNGDCSTAL-VGGVAVL-----ATPGAFIEFS 10497

```

QY 338 -----RNVGHMTNDAIVDTGSEV--FEGIMDALFTG--LIAIHGKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLAWGEGVAVLLERLSDRKGRHVLAV-----VRGSA 10550
QY 385 INSTGSIYIVKPKKHGPAEVAFTCELSRVEDVLGPPQNTMKIGIMDEERRTTVNLKAC 444
Db 10551 INQDGASNGLTAP--HGFSQ-----OHLIRQALAD--AELT----- 10582
QY 445 IKAADRVVFINTEGFLDRTGDEIHT-----SMEAGPMVRKGTWMSQ----- 485
Db 10583 ---SSDVVEGHTGTGLGPIEAQALLATYGGCRAPGQPLRLGLTUKSNIGHTQAASGV 10639
QY 486 -----PWLAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKMQALRHGVLPKTLHVDEPTQDQVNSAGSVELLTEAVDWPGRRLRAGVSAGF 10699
QY 509 QVKGKMTWTELMADWVETKIAQPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEBAPAEVSPAVEPPAGGVVFPVPSAKTSAALDAQIGQLAAVAEDRTDV 10759
QY 558 DVAAVQOGL-----AGKRRTATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAVGDSREALRDLRMPEGLVRGTVTDPGRVAVFPQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVVDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAELDSSPEFAAAMAECELTALSPYVDWSLEAVVRQAPSAPTLDRLVDVQVP-VTF 10878
QY 637 RISSOLLANWLRHGVTISADVRASLERMAPLVDRQAGDVAYRPMAPNFDSDS 688
Db 10879 AVMSVLAKVQHHGHTPEAVTGHSGGEIA-----AAYVAGALTLLDA 10920

```

RESULT 13

```

US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

```

```

Query Match 3.4%; Score 128; DB 10; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIRAVLYDFVNNALPCTDIDPDSFWAGVDKVVADLTTPQNALNARDELQACIDKWHR 68
Db 10193 LRDAVGL-----DTVLRLTGTETPEPGSGDGAADFGAEPFEASL---DLDLDAEA---LI 10241
QY 69 RRVTEPID-----MDAYRQFLTEIGVLLPEPDDFTITTSVGDAEITTTAGPQLVWP 119
Db 10242 RMAIGPRTNTSSNEQLVDALRASLKENEELEKES-----RRRDRPREPMAIVG 10291
QY 120 VLNARFALNAANRWGSLYDAL-YGTDV---IPTDGAEXGPTNVKVRGDKVYAYARK-- 173
Db 10292 -MSCRFAGGIRSP--DLMDAAVAGKDLVSEVPEERGWGIDSLYDPVFGKRGKTTYVRNAA 10348

```

```

QY 174 FLDDSPVSSSFG-----DATGFTVQDQQLVVALPDKSTGLANPQGFAGYTGAAESTSV 229
Db 10349 FLDDAAGFAAAGGISTPREALAMPQQRQLLEA-----SWEVFERAGIDPASV 10396
QY 230 LLINHGLHI-----EILIDPESQVGTTRAGVKDVILESAITIMDFEDSVAADA 280
Db 10397 RGTGVTGVVGGCYQYADIRVAPEGTGGYV-VTGNSSAVASGRIAYSGLGEGPAVTVDT 10455
QY 281 A---DKVLGYZNWLGKGLAAAVDKDGTAFRLVLRDRNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHLALKGLRNGDCSTAL--VGVVAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHMTNDAIVDTGSEV--FEGIMDALFTG--LIAIHGKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLAWGEGVAVLLERLSDRKGRHVLAV-----VRGSA 10550
QY 385 INSRGTGSIYIVKPKKHGPAEVAFTCELSRVEDVLGPPQNTMKIGIMDEERRTTVNLKAC 444
Db 10551 INQDGASNGLTAP--HGFSQ-----OHLIRQALAD--AELT----- 10582
QY 445 IKAADRVVFINTEGFLDRTGDEIHT-----SMEAGPMVRKGTWMSQ----- 485
Db 10583 ---SSDVVEGHTGTGLGPIEAQALLATYGGCRAPGQPLRLGLTUKSNIGHTQAASGV 10639
QY 486 -----PWLAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKMQALRHGVLPKTLHVDEPTQDQVNSAGSVELLTEAVDWPGRRLRAGVSAGF 10699
QY 509 QVKGKMTWTELMADWVETKIAQPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEBAPAEVSPAVEPPAGGVVFPVPSAKTSAALDAQIGQLAAVAEDRTDV 10759
QY 558 DVAAVQOGL-----AGKRRTATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAVGDSREALRDLRMPEGLVRGTVTDPGRVAVFPQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRVVDQGV--GCSKVPDIHDVVALMEDRATL 636
Db 10820 QWAGMGAELDSSPEFAAAMAECELTALSPYVDWSLEAVVRQAPSAPTLDRLVDVQVP-VTF 10878
QY 637 RISSOLLANWLRHGVTISADVRASLERMAPLVDRQAGDVAYRPMAPNFDSDS 688
Db 10879 AVMSVLAKVQHHGHTPEAVTGHSGGEIA-----AAYVAGALTLLDA 10920

```

RESULT 14

```

US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-6

```

```

Query Match 3.4%; Score 128; DB 11; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIRAVLYDFVNNALPCTDIDPDSFWAGVDKVVADLTTPQNALNARDELQACIDKWHR 68

```

10193 LRDAVGL-----DIVLRLTGIEPEPGSGSDGGAADPGAEPEASI---DDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVVP 119
Db 10242 RMAIGPRNTMTSSNEQLVDALRASLKENEELRKES-----RRRADRQEPMAIVG 10291
QY 120 VLNAAREALNANARWGSILYDAL-YGTDV---IPETDGAEGKPTYNKVRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSPB--DLMDAVAAGKDLVSEVPEERGWIDSLYDPVGRKGTYYVRNAA 10348
QY 174 FLDSVPLSGSGFG---DATGFTVQDGLWALPKSTGLANPGQFAGYTGAAESPTSV 229
Db 10349 FLDDAAGFADAFGISPREALAMPQORQLLEA-----SWEFERAGIDPASV 10396
QY 230 LLINHGHLI-----ELIDPESQVGTDRAGVKDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTDVGVVCGGYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGRISAYSLGLEGPATVDT 10455
QY 281 A---DKVLGYRNWLGKDLAAAVKDGTAFLRVNLNRDNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHKLKGLNGDCSTAL-VGGYAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHLMTNDAIVDTDGSEV--FEGIMDALFTG--LIAIHGLKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLANGEGVAVLLERLSDARRKGRVLA--V-----VRGSA 10550
QY 385 INSRGTSIYIVKPKWGHGPAEVAFTCELSRVEDVLGPONTWKIGIMDEERTTNLAKAC 444
Db 10551 INQDASNGLTAP--HGFSQ-----QHLIRQALAD--ARLT----- 10582
QY 445 IKAADRVVFIINTGFLDRTGDEIHT-----SMEAGPMVRKGMKSO----- 485
Db 10583 ---SSDVEDVVEGHGTGRLGDPTEAQALLATYQGRAPQPLRLGLTKSNIGHTQAAAGV 10639
QY 486 -----PWILAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKQVQALRHGVLPKTLHVDEPTQDQVDSAGSVELLTEAVDWPFRPGLRRAGVSAGF 10699
QY 509 QVKGKMTMTLMADWVETKIAOPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEEAFAVEEPAGGVVPPVPSAKTSAALDAQIGQLAAAYADRTDV 10759
QY 558 DVAAVQOGL-----AGKRATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAGVDSREALDRALRMEPEGLVRGTVPDGRVAFVFGQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRWVDCGV--GCSKVPDIHDVLMEDRATL 636
Db 10820 QWAGMGAEILDSPEFAAAVACETALSFPYDWSLEAVVROAPSAPFTLDRVDVQVP-VTF 10878
QY 637 RISSQLLANWLRHGVTISADVRSASLERMAPLVDQRQAGDVAYVRPAPNPFDD 688
Db 10879 AVMVSLAKVQHHGITPEAVIGHSGGEIA-----AAVAGALTLLDDA 10920

RESULT 15

US-10-271-889-49
; Sequence 49, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600 582U1
; CURRENT APPLICATION NUMBER: US/10/271,889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-271-889-49

Query Match 3.4%; Score 128; DB 12; Length 11877;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 172; Conservative 85; Mismatches 319; Indels 256; Gaps 40;

QY 9 LRIRLVYDFVNNALPGTDIDPDSFWAGVVKVADLTPQNALNARDELQAOIKDKWHR 68
Db 10193 LRDAVGL-----DIVLRLTGIEPEPGSGSDGGAADPGAEPEASI---DDLDAEA---LI 10241
QY 69 RRVIEPID-----MDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVVP 119
Db 10242 RMAIGPRNTMTSSNEQLVDALRASLKENEELRKES-----RRRADRQEPMAIVG 10291
QY 120 VLNAAREALNANARWGSILYDAL-YGTDV---IPETDGAEGKPTYNKVRGDKVIAYARK-- 173
Db 10292 -MSCRFAGGIRSPB--DLMDAVAAGKDLVSEVPEERGWIDSLYDPVGRKGTYYVRNAA 10348
QY 174 FLDSVPLSGSGFG---DATGFTVQDGLWALPKSTGLANPGQFAGYTGAAESPTSV 229
Db 10349 FLDDAAGFADAFGISPREALAMPQORQLLEA-----SWEFERAGIDPASV 10396
QY 230 LLINHGHLI-----ELIDPESQVGTDRAGVKDVILESAITTIMDFEDSVAADA 280
Db 10397 RGTDVGVVCGGYQDYAPDIRVAPEGTGGYV-VTGNSSAVASGRISAYSLGLEGPATVDT 10455
QY 281 A---DKVLGYRNWLGKDLAAAVKDGTAFLRVNLNRDNYTAPGGGQFTLPGRSLMFV 337
Db 10456 ACSSSLVALHKLKGLNGDCSTAL-VGGYAVL-----ATPGAFIEFS 10497
QY 338 -----RNVGHLMTNDAIVDTDGSEV--FEGIMDALFTG--LIAIHGLKASDVNGPL 384
Db 10498 SQQAAADGRTKGFASAADGLANGEGVAVLLERLSDARRKGRVLA--V-----VRGSA 10550
QY 385 INSRGTSIYIVKPKWGHGPAEVAFTCELSRVEDVLGPONTWKIGIMDEERTTNLAKAC 444
Db 10551 INQDASNGLTAP--HGFSQ-----QHLIRQALAD--ARLT----- 10582
QY 445 IKAADRVVFIINTGFLDRTGDEIHT-----SMEAGPMVRKGMKSO----- 485
Db 10583 ---SSDVEDVVEGHGTGRLGDPTEAQALLATYQGRAPQPLRLGLTKSNIGHTQAAAGV 10639
QY 486 -----PWILAYED--HNVD--AGLA-----AGFSGRA 508
Db 10640 AGVIKQVQALRHGVLPKTLHVDEPTQDQVDSAGSVELLTEAVDWPFRPGLRRAGVSAGF 10699
QY 509 QVKGKMTMTLMADWVETKIAOPRAGASTA-W-VPSPTAATL-----HALHYHQV 557
Db 10700 VGGTNAHVLEEAFAVEEPAGGVVPPVPSAKTSAALDAQIGQLAAAYADRTDV 10759
QY 558 DVAAVQOGL-----AGKRATIEQLLTIP--LAKELAWAPDEI----- 593
Db 10760 DPAVAARALVDSRTAMEHRAVAGVDSREALDRALRMEPEGLVRGTVPDGRVAFVFGQGT 10819
QY 594 -----REEVDNN-----COSILGYVVRWVDCGV--GCSKVPDIHDVLMEDRATL 636
Db 10820 QWAGMGAEILDSPEFAAAVACETALSFPYDWSLEAVVROAPSAPFTLDRVDVQVP-VTF 10878
QY 637 RISSQLLANWLRHGVTISADVRSASLERMAPLVDQRQAGDVAYVRPAPNPFDD 688
Db 10879 AVMVSLAKVQHHGITPEAVIGHSGGEIA-----AAVAGALTLLDDA 10920

Mon Nov 24 15:01:42 2003

us-09-688-672a-2.rapb

Page 10

Search completed: November 21, 2003, 16:38:09
Job time : 32.8617 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 15.4376 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-2
Perfect score: 3810
Sequence: 1 TDRVSVGNLRIRARVLYDFVN.....KARAAEKPAAPSRAAGDRAAR 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:**

1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	741	2 F70722	probable glcB prot
2	3117.5	81.8	731	2 T44752	probable malate sy
3	2569	67.4	725	2 H83586	malate synthase G
4	2374.5	62.3	744	2 AH2582	malate synthase G
5	2374.5	62.3	744	2 F97364	malate synthase G
6	2318	60.8	728	2 AF3299	malate synthase (E
7	2289	60.1	739	2 I40715	malate synthase (E
8	2256.5	59.2	727	2 E83916	malate synthase BH
9	2154.5	56.5	723	2 S51788	malate synthase (E
10	205.5	5.4	542	2 S17774	malate synthase (E
11	190.5	5.0	533	2 D91245	malate synthase A
12	190.5	5.0	533	2 B86093	malate synthase A
13	190	5.0	533	2 AD1011	malate synthase (E
14	189.5	5.0	533	1 SYECMA	malate synthase (E
15	187	4.9	556	2 H82287	malate synthase A
16	186	4.9	551	2 JX20196	malate synthase (E
17	185	4.9	551	2 JX20196	malate synthase (E
18	175.5	4.6	525	2 A87468	malate synthase (E
19	175	4.6	528	2 S17773	malate synthase (E
20	171.5	4.5	559	2 T03412	malate synthase (E
21	165	4.3	554	2 S26645	malate synthase (E
22	165	4.3	554	2 S48493	malate synthase (E
23	160	4.2	555	1 SYHQMA	malate synthase (E
24	160	4.2	562	2 T48413	malate synthase-l1
25	154.5	4.1	562	2 E82396	malate synthase-re
26	151.5	4.0	564	2 T07690	malate synthase (E
27	150.5	4.0	561	1 SYRPM	malate synthase (E
28	149.5	3.9	566	2 S44186	malate synthase (E
29	149	3.9	532	2 AF0453	malate synthase (E

30 149 3.9 566 2 S15387
31 144.5 3.8 568 1 SYKWA
32 139 3.6 3455 2 B82519
33 138 3.6 567 1 SYCSM2
34 137.5 3.6 4848 2 T30289
35 135.5 3.6 521 2 D75581
36 135.5 3.6 3442 2 B82589
37 134.5 3.5 1441 2 B86807
38 130 3.4 1624 2 C70867
39 129.5 3.4 597 2 A90770
40 129.5 3.4 597 2 B85632
41 129 3.4 5188 2 B85547
42 129 3.4 5291 2 F90696
43 128.5 3.4 597 1 HQECL
44 128.5 3.4 808 2 F70720
45 127.5 3.3 26926 1 I38344

ALIGNMENTS

RESULT 1

F70722

probable glcB protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70722

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Article: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70722

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 <COL>

A:Cross-references: GB:278020; GB:AL123456; NID:93261625; PIDN:CAB01465.1; PID:e257679;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: glcB

Query Match	100.0%	Score 3810	DB 2	Length 741
Best Local Similarity	100.0%	Pred. No. 3.7e-246	Mismatches 0	Indels 0
Matches 740	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	TDRVSVGNLRIRARVLYDFVNNEALPGTDIDPDSFWAGYDKVAVDLTPQNQALLNARDELQ	60	
Db	2	TDRVSVGNLRIRARVLYDFVNNEALPGTDIDPDSFWAGYDKVAVDLTPQNQALLNARDELQ	61	
QY	61	AQIDKWHRRRVIEIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAITTTAGQLVVPV	120	
Db	62	AQIDKWHRRRVIEIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAITTTAGQLVVPV	121	
QY	121	LNARFALNANARNGSLYDALYGTDPVETDGAEGKGTYNKVRGDKVIAYARKFLDSDVP	180	
Db	122	LNARFALNANARNGSLYDALYGTDPVETDGAEGKGTYNKVRGDKVIAYARKFLDSDVP	181	
QY	181	LSSGSFGDATGFTVQDQGVVALPKSTGLANPQCFAGYTGAAESPTSVLLINHLHIEI	240	
Db	182	LSSGSFGDATGFTVQDQGVVALPKSTGLANPQCFAGYTGAAESPTSVLLINHLHIEI	241	
QY	241	LIDPESQVGTTRAGVKDVILESAITTIMDPEDSVAADKVLGVRNGLNKGDLAA	300	
Db	242	LIDPESQVGTTRAGVKDVILESAITTIMDPEDSVAADKVLGVRNGLNKGDLAA	301	
QY	301	AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGSRSLMFRVNVGHLMTNDAIVDTDGSEVFEG	360	
Db	302	AVDXDGTAFRLVLRNDRNYTAPGGQFTLPGSRSLMFRVNVGHLMTNDAIVDTDGSEVFEG	361	
QY	361	IMDLFTGLIAHGLKASDVNGPLNSRTGSIITVKEPMHGPAPVATCLFSRVEDVLG	420	

Db 362 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSFVEDVLG 421
Qy 421 LPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRKG 480
Db 422 LPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGPVVRKG 481
Qy 481 TMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAQPRAGASTAW 540
Db 482 TMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAQPRAGASTAW 541
Qy 541 VPSPTAATLHALHYQVDVAQVQOGLAGKSRATIEQLLTIPLAKELAWADEIREEDNN 600
Db 542 VPSPTAATLHALHYQVDVAQVQOGLAGKSRATIEQLLTIPLAKELAWADEIREEDNN 601
Qy 601 CQSILGYVVRVWDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSA 660
Db 602 CQSILGYVVRVWDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSA 661
Qy 661 LERMAPLVDRQVAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPIHLRRREF 720
Db 662 LERMAPLVDRQVAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPIHLRRREF 721
Qy 721 KARAAEKPAQSDRAGDDAAR 740
Db 722 KARAAEKPAQSDRAGDDAAR 741

RESULT 2

T44752
probable malate synthase (EC 4.1.3.2) G [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
C:Accession: T44752
R:Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z22833
A:Accession: T44752
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-731 <PAR>
A:Cross-references: EMBL:AL008609; PIDN:CAA15459.1
A:Experimental source: cosmid B1788
C:Genetics:
A:Note: glcB
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 81.8%; Score 3117.5; DB 2; Length 731;
Best Local Similarity 82.0%; Pred. No. 5.7e-200;
Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;

Qy 1 TDRVSVGNLRIRARLVYDFVNNALPGTDIDPDSEWAGVDKVVADLTPOQALLNARDLQ 60
Db 2 TDRVSVGNLRIRARLVYDFVNNALPGTDIDPDSEWAGVDKVVADLTPOQALLNARDLQ 61
Qy 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 120
Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 121
Qy 121 LNARFALNANARWGLSDYALYGTDTVETPGASKGPTYNKVRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNANARWGLSDYALYGTDTVETPGASKGPTYNKVRGDKVIAYARKFLDSDVP 181
Qy 181 LSSGSFGDATGFTVDDQGLVVALPDKSTGLANPQFAGYTGAAESPT-SVLLINHLHIE 239
Db 182 LASDSWTNATGVSIPDGLQIAIGTNSGLASPEKFGVYNNRQLRSSNWSVLLANHLHIE 241
Qy 240 ILIDPESQVGTTRAGVNDVILESAITTIMDFESVAAVDAADKVLGVNWLGNKGLA 299
Db 242 VLIDPESPIGKTPVGINKDVIILESAITTIMDFESVAAVDAADKVLGVNWLGNKGLD 301
Qy 300 AAVDKDGTAFRLNDRNRYTAPGGQFTPLGRSLMFRVNYGHLMTNDIAV- ---DTDGS 355
Db 302 BEVNKDGKTFTRVLNADRSYTPDGGELTLPGRSLLFVRNVLHGLTTSDAILDVGGDQGEK 361

Qy 356 EVFEGIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSFV 415
Db 362 EVFEGIIDAVFTGLAAIHGLTGEANGPLTNSRTGSIYIVKPKMHGPAEVAFTCELSFV 421
Qy 416 EDVLGLPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 475
Db 422 EDVLGLPONTMKGIMDEBERTTVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMEAGP 481
Qy 476 MYRKTMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAQPRAG 535
Db 482 MYRKTMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTMELMADWVETKIAQPRAG 541
Qy 536 ASTAWPSPPTAATLHALHYQVDVAQVQOGLAGKSRATIEQLLTIPLAKELAWADEIRE 595
Db 542 ASTAWPSPPTAATLHALHYQVDVAQVQOGLAGKSRATIEQLLTIPLAKELAWADEIRE 601
Qy 596 EVDNCCQILGYVVRVWDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSA 655
Db 602 EVDNCCQILGYVVRVWDQGVGCSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSA 661
Qy 656 DVRASLERMAPLVDRQVAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPIHLR 715
Db 662 DVRASLERMAPLVDRQVAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPIHLR 721
Qy 716 RRRREFKAR 723
Db 722 RRRREFKAR 729

RESULT 3

H83586

malate synthase G PA0482 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83586
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83586
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <STO>
A:Cross-references: GB:AE004485; GB:AE004091; NID:g9946332; PIDN:AAG03871.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: glcB; PA0482

Query Match 67.4%; Score 2569; DB 2; Length 725;
Best Local Similarity 69.1%; Pred. No. 2.2e-163;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

Qy 1 TDRVSVGNLRIRARLVYDFVNNALPGTDIDPDSEWAGVDKVVADLTPOQALLNARDLQ 60
Db 2 TDRVSVGNLRIRARLVYDFVNNALPGTDIDPDSEWAGVDKVVADLTPOQALLNARDLQ 61
Qy 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 120
Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 121
Qy 121 LNARFALNANARWGLSDYALYGTDTVETPGASKGPTYNKVRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNANARWGLSDYALYGTDTVETPGASKGPTYNKVRGDKVIAYARKFLDSDVP 181
Qy 181 LSSGSFGDATGFTVDDQGLVVALPDKS-TGLANPQFAGYTGAAESPT-SVLLINHLHIE 239
Db 182 LSSGSFGDATGFTVDDQGLVVALPDKS-TGLANPQFAGYTGAAESPT-SVLLINHLHIE 241
Qy 240 ILIDPESQVGTTRAGVNDVILESAITTIMDFESVAAVDAADKVLGVNWLGNKGLA 299

Db 242 IOTDPPSPVQGTDAAGVKDVLMEALTTIMDCEDSVAAVDADDKVVIVRNWMLGMLKGDIA 301
QY 300 AAVDKDGTAFVLARDRNYTAPGGQFTLPGRSLMEVENVGHMTNDALVDTDCSEVFE 359
Db 302 EVSKGSGSTFTTMDPDRVYTTADGSELTLGRSLJFVRNVLHMTNDAILDKDGNVFE 361
QY 360 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKVHGAFAVFTCELSERVED 417
Db 362 GICDGLFTSLIAIH-----DLNGNTRSKNSRTGSIYIVKPKVHGPPEAAFTNELFGRVED 416
QY 418 VLGLPONTMKGIMDEBERTTNLAKACIAAADRVVFNITGLDRTGDBEIHSTMEAGPMV 477
Db 417 VLGLPNTLUKVGIMDEBERTTNLAKACIAAADRVVFNITGLDRTGDBEIHSTMEAGAV 476
QY 478 RKGTKSQPWILAYEDHNVDAAGLAFSGRAQVGVKGMWMTMELMADMVETKIAQPRAGAS 537
Db 477 RKGAMKSEKWIYAYENNVDVGLATGLQKQAKIIGKGMWMPDLMAMLEQKIGHPLAGAN 536
QY 538 TAWVPSPTAATHALHYHOVDVAAVQOGLAGKRRATIEBOLLTIPLAKELANAPDIREV 597
Db 537 TAWVPSPTAATHALHYKVDVFAQAEAKRTPASVDDILTIPLAPNTNNTWAEIKREV 596
QY 598 DNNQSGILGVVVRWDQVGVCSKVPDIHDVLMEDRATLRISQALLANLHGVITSADV 657
Db 597 DNNAGILGVVVRWDQVGVCSKVPDINDVGLMEDRATLRISQALLANLHGVISQEQV 656
QY 658 RASLERMAPLVRQNAVGVAYRPMAPNFDDSIAPLAAQELILSGAQPNNGYTEPILHRRR 717
Db 657 VESLKEMAVVDRQNASDPSYRPMAPNFDDNVAFQAALELVVEGTRQPNNGYTEPVLHRRR 716
QY 718 REFKA 723
Db 717 REFKA 722

RESULT 4
AH2582
malate synthase G [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2582
R:Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AH2582
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41078.1; PID:g17738367; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glcB
A:Map position: circular chromosome

Query Match 62.3%; Score 2374.5; DB 2; Length 744;
Best Local Similarity 62.3%; Pred. No. 2.1e-150;
Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;

QY 1 TDRVSVGNL-----RIARV-----LYDFVNEALPGTDIDPDSEWAGVDKVA 43
Db 6 SDRLSVQSPSEYKEAHVSRDTKFLGSLDRLVAFLTDEVLPGTGLDSEITFEFGSAIVH 65
QY 44 DLTPOQALLNARDELQADIKWHRRVIEPDMDAYQFLTEIGYLLPEDDFTTTSG 103
Db 66 ELSKPNRELLAKDALQEKIDGWRENG-APSDFOAYEAFLEKEIGYLLPEGPGFKVETNN 124
QY 104 VDAEITTTAGPOLVPLNARALNANRWGSLYDALYGTDIVETPGAEGKPTYNKR 163

Db 125 VDPEIAVAGPQLVVPVMNARYALNANRWGSLYDALYGTDAISDADGAEGKRGYNPKR 184
QY 164 GDKVIAVARFELDDSVPLSGSGFDATFTVQDQQLVVALPDKSTGLANPQAGYTGAA 223
Db 185 GDKVIAWARFLDESAPLETGSDVDTGFNIADGLQLAIGAAATGLKDAVQVQFSGSEA 244
QY 224 ESPTSVLLIINHGLHIEILIDPESOVGTTDRAGVKDVILESAITTIMDFESVAAVDAADK 283
Db 245 AKPATILLKGNLHTEIVIDPSTEIGKSDRAGISDVILESALTITIMDCESVAAVDAEDK 304
QY 284 VLGRVNLGNKGLDAAAADKGTAFULVLRNRDNYTAPGGQFTLPGRSLMFVRNVGHL 343
Db 305 VLVTGNVLGMRGLTEAVSGKGNITFTRLNPDYRYTAPDGSALTLPGRSLMLVRNVGHL 364
QY 344 MTNDALVDTDGSVEFEGIMDALFTGLTAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPA 403
Db 365 MTNPAILDRGRVPEGIMDAVVTALLALYDVGPS---GRQNSRAGSMVTVFVKMHGPE 421
QY 404 EVAFTCLFERVEDVLGPQNTMKIGIMDEBERTTNLAKACIAAADRVVFNITGLDRT 463
Db 422 EVAFANEIFARVENLVGMAFNTMKGIMDEBERTTNLAKESIRAAKDRVVFNITGLDRT 481
QY 464 GDEHTSWEAGPVRKCTKMSQPMILAYEDHNVDAAGLAFSGRAQVGVKGMWMTMELMAD 523
Db 482 GDEHTSWEAGPMIRKGMKQAAIAAYENNVNDIGLECSLGHQAQIGKGMWMPDLMAA 541
QY 524 MVETKIAQPRAGASTAVWPSPPTAATHALHYHOVDVAAVQOGLAGKRRATIEOLLTIPLA 583
Db 542 MLEQKIAHPKAGANTAVWPSPPTAATHALHYKVDVAAVQOGLKSRGAKLSDLISVPVA 601
QY 584 KELAWAPDEIREVEVDNNQCSILGVVVRWDQVGVCSKVPDIHDVLMEDRATLRISQALL 643
Db 602 PRPNWTEETQRELDNNAQILGVVVRWDQVGVCSKVPDINNIGLMEDETRATLRISAQM 661
QY 644 ANWLHRGVITSADVRASLERMAPLVRQNAVGVAYRPMAPNFDDSIAPLAAQELILSGAQ 703
Db 662 ANWLHRGVITAEQIITKMKMAAVVDTONAGDPAFLPMASDFDGSVAFQAAVELVLKRE 721
QY 704 QPNGYTEPILHRRRREFKARAA 725
Db 722 QPNGYTEPILHRRRLELKAQA 743

RESULT 5
F97364
malate synthase G (PA0482) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: F97364
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85871.1; PID:g15154912; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C78
A:Map position: circular chromosome

Query Match 62.3%; Score 2374.5; DB 2; Length 744;
Best Local Similarity 62.3%; Pred. No. 2.1e-150;
Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;

QY 1 TDRVSVGNL-----RIARV-----LYDFVNEALPGTDIDPDSEWAGVDKVA 43
Db 6 SDRLSVQSPSEYKEAHVSRDTKFLGSLDRLVAFLTDEVLPGTGLDSEITFEFGSAIVH 65
QY 44 DLTPOQALLNARDELQADIKWHRRVIEPDMDAYQFLTEIGYLLPEDDFTTTSG 103

Db 66 ELSPQNRRELLAKRDLQEKIDGWYRENG-APSPDFDAYEAFLEKIGYLLPEGPGFKVETNN 124
QY 104 VDAEITTTAGPQLVVPVNLNARFALNANAGSLYDALYGTVDVTPETDGAEGKPTNKVR 163
Db 125 VDEIAVAGPQLVVPVNNARYALNANAGSLYDALYGTDAISADGAEGKRGYNPKR 184
QY 164 GDKVIARPLDSDVPLSSGSGDGTFTVQDQQLVVALPKDSTGLANPGQAGVTGA 223
Db 185 GDKVIARWFLDSDAPLETGSDVTFGNFIADQLLQALGAATTLGLKAVQKFGSGEA 244
QY 224 ESPTSLLIHGHGHIETLIDPEQVGTTRAGVKDVIKESAITTINDFDSVAADK 283
Db 245 AKPATILGKNGHLTEIVDPSTEIGKSDRAGISDVIKESAITTINDFDSVAADK 304
QY 284 VLGVRNMLGKNGDLAAADKDTAFVLNRDRNTYAPGGGFTLPGSLMFRVNVHL 343
Db 305 VLVYGNMGLMRGLTEAVSKGNTFTFRLNDRNTYAPDGSALTLPGRSLMLVRNVHL 364
QY 344 MTNDALVDTGSEVFEIGMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPA 403
Db 365 MTNPAILDRDGRDVEGIMDAVVTALIALYDVGPS---GRRQNSRAGSMFVVRKPHGPE 421
QY 404 EVAFTCLFSEVDEVLGTPQNTKIGIMDEERTTNLAKICIAAADRVVVFINTGLDRT 463
Db 422 EVAFAEIFARVENLVGMATNWKMGIMDEERTTNLAKESIRAAKDRVVVFINTGLDRT 481
QY 464 GDEIHTSMEAGPMVRKGTMSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTWELMAD 523
Db 482 GDEIHTSMEAGPMVRKGTMSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTWELMAD 541
QY 524 MVEKIAQPRAGASTAVPSPSTAATHALYHVDVAAGVQGLAGKRATIEOLLTIPIA 583
Db 542 MLEOKIAHPKACANTAVPSPSTAATHALYHVDVAAGVQGLAGKRATIEOLLTIPIA 601
QY 584 KELAWAPDEIRREVDNNQCSILGVYVRWDQGVCSKVPDIHVALMEDRATIRISQLL 643
Db 602 PRPNWTPEEIQRELDNNAQILGVYVRWDQGVCSKVPDIHVALMEDRATIRISQAQM 661
QY 644 ANWLHGVITTSADYRASIEMAPLVDRQAGDVAYRPMAPNFDSDTAFLAAQELISGAQ 703
Db 662 ANWLHGVITTSADYRASIEMAPLVDRQAGDVAYRPMAPNFDSDTAFLAAQELISGAQ 721
QY 704 QPNGYTPILHRRRRREFKARAA 725
Db 722 QPNGYTPILHRRRRREFKARAA 743

RESULT 6

AF3299
malate synthase (EC 4.1.3.2) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AF3299
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Eizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <KUR>
A:Cross-references: GB:AB008917; PIDN:AAL51561.1; PID:gl7982281; GSFDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0380
A:Map position: 1
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 60.8%; Score 2318; DB 2; Length 728;
Best Local Similarity 61.8%; Pred. No. 1.2e-146;
Matches 444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;

QY 4 VSVGNLRIRARVLYDFVNNALPGTDDIPDSFWAGVDKVVADLTLPONQALLNARDELQAOI 63
Db 10 VEIEGLAVAPLEVEFLAKERAPGTGVEPEKFWKGAFAIIRDLAPKRNALLAKRDELQARI 69
QY 64 DKWRRRRVIEPDMDAYRQFLTEIGYLLPPDPDDFTTTTSGVDAEITTTAGPQLVVPVYNA 123
Db 70 DAWKENRDKYSQADVQOFLKIDIGYLLPEGASVSTTNVDPETHIAGPQLVVPVYNA 129
QY 124 RFALNANARWGSYDALYGTVDVTPETDGAEGKPTNKVRGDKVIAVARFLDSDVPLSS 183
Db 130 RYALNANARWGSYDALYGTDAISEADGAEGKGYNPARGKEKVIKAWKFLDSEAPLST 189
QY 184 GSGDGTCTFTVQDQQLVVALPKDS-TGLANPGQAGYTGAAESPTSVLLNHGHGHIILI 242
Db 190 GKWADVAGLANVKGLEIRLTGDSATTLKDSQFKGYNGDAASPTNVLLAKHNVHVDIVI 249
QY 243 DPESQVGTTRAGVKDVIKESAITTINDFDSVAADKAVLGYRNWGLNKGDLAAAV 302
Db 250 NADHPIGKTDPAHIADVVLKESAITTINDFDSVAADKAVLGYRNWGLNKGDLAAAV 309
QY 303 DKDGTAFVLNRDRNTYAPGGGFTLPGSLMFRVNVHLMTNDALVDTGDSVFEIGM 362
Db 310 EKNGQWTRRLNGDRTYTPDGSSTLTLLKGRSLMLVRNVGHLMNTNPAILDAGNEVPEGIM 369
QY 363 DALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLP 422
Db 370 DAAFTSLIALHDI---GNGRHNRSRGSVVIVKPKHGPPEVAFANEFTFTEMLGMK 426
QY 423 QNTMKIGIMDEERTTNLAKICIAAADRVVVFINTGLDRTGDSIHTSMEAGPMVRKGTM 482
Db 427 PNTLIGIMDEERTTNLAKESIRAAKDRVVVFINTGLDRTGDSIHTSMEAGPMVRKGTM 486
QY 483 KSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTWELMADMTETKIAQPRAGASTAVP 542
Db 487 KQANWIGAYEOWNVDIGLECSLGHQAIGKGNWAMPDMWAAMLEQKTAHPKAGANTAVP 546
QY 543 SPTAATLHALYHVDVAAGVQGLAGKRATIEOLLTIPIAKELAWAPDEIRREVDNNQ 602
Db 547 SPTAATLHALYHVDVAAGVQGLAGKRATIEOLLTIPIAKELAWAPDEIRREVDNNQ 606
QY 603 SILGVYVRWDQGVCSKVPDIHVALMEDRATIRISQLLNLHGHGHIILISADYRASI 662
Db 607 GLGVYVRWDQGVCSKVPDIHVALMEDRATIRISQAQIANWLYHGVVSEAGVMEATM 666
QY 663 RMAPLVDRQAGDVAYRPMAPNFDSDTAFLAAQELISGAQVQNGYTPILHRRRRREFK 721
Db 667 RMAAIVDQKNEGDPYRPMAPNFDSDTAFLAAQELISGAQVQNGYTPILHRRRRREFK 725

RESULT 7

I40715
malate synthase (EC 4.1.3.2) - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-May-2000
C:Accession: I40715; I40836
R:Reinisch, D.J.; Eikmanns, B.J.; Sahm, H.
Microbiology 140, 3099-3108, 1994
A:Title: Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene
A:Reference number: I40715; MUID:951111631; PMID:7812449
A:Accession: I40715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: EMBL:X78491; NID:9530011; PIDN:CAA55243.1; PID:9530012
R:Lee, H.
J. Microbiol. Biotechnol. 4, 256-263, 1994
A:Title: Molecular characterization of aceB, a gene encoding malate synthase in Coryneb
A:Reference number: I40836
A:Accession: I40836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: GB:L27123; NID:9653573; PIDN:AAA68074.1; PID:9853574

[illegible]

```
Db          723 AK6 725
::
C.Species: Escherichia coli (strain K-12)
C.Date: 23-Aug-1995 #sequence_revision 13-Mar-1997 #text_change 01-Mar-2002
C.Accession: S51788; F65083
R.Molina, I.; Pellicer, M.T.; Badia, J.; Agular, J.; Baldoma, L.
Eur. J. Biochem. 224, 541-548, 1994
A.Title: Molecular characterization of Escherichia coli malate synthase G. Differentiat
```

A:Reference number: S51788; MUID:95010032; PMID:7925370

A:Accession: S51788

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-723 <MO>

A:Cross-references: EMBL:X74547; NID:9517246; PIDN:CAA52639.1; PID:9517247

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F65083

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-723 <BLAT>

A:Cross-references: GB:AE000380; GB:U00096; NID:91789344; PIDN:AAC76012.1; PID:91789348

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: glcB

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 56.5%; Score 2154.5; DB 2; Length 723;

Best Local Similarity 58.0%; Pred. No. 9.7e-136;

Matches 419; Conservative 115; Mismatches 174; Indels 15; Gaps 5;

QY 4 VSVGNLRIRVLYDFVNEALPGTDIDPDSFMAGVQKVAADLPONCALNARDELQAQI 63

DB 5 ITQRLRIDANFKFVDEEVLPGLDAAAFWRNFDIIVHDLAPENQLLAEDRQAAL 64

QY 64 DKWHRRIEVIDMDAYRQFTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPVNA 123

DB 65 DEWHRSPGPVKDAAKYSFLRELGLVLPQPERVTVETTGIDSEITTSQAGPQLVWPVNA 124

QY 124 RFLANANARAGSLYDALYDVTDPEDGAEGKTYNKVRGDKVIARFELDDSVPLSS 193

DB 125 RYALNANARAGSLYDALYDGSIIIPQEGAMVSG--YDPQGEQVIAWRFLDESUPLEN 192

QY 184 GSFQDGTFTVQDQQLVVALPD-KSTGLANPGQFAGYTGAAESPTSVLLINHGHIETILI 242

DB 183 GSYQDVAFKVVQKQAIQLKNGKETTLLTPAQVGVYRGDAAPTCLLKNNGHIELOI 242

QY 243 DPESQVTTDRAGVKDVLISAITTMDFEDSVAAVDAADKVLGYRNWLGKNGDLAAAV 302

DB 243 DANGRIKDDPAHNDIVIAAISTILDCEDSVAAVDAEDKILLRYNLLGLMQGTLQKRM 302

QY 303 DKDGTAFRLVLRDRNTYAPGGGQFTLPGRSLMFVRNVGHMTNDATVDTDGSSEVFEGIM 362

DB 303 EKNGRQIVKLNDRPHYTAADGSIISLHGRSLFIRNVGHMTIPVINDSEGNEIPEGIL 362

QY 363 DALFTGLIAIHLKASDVNGPLINSRTGSIYIVKPKHGPFAVFTCELSRVEDVLGLP 422

DB 363 DGVMTGAIALYDLKVQK-----NSRTGSIYIVKPKHGPQEVAFANKLFTRIETMLGMA 416

QY 423 QNTWKIGTMDERTTNNLKACIAAADRVVFINTEGFLDRTGDEIHTSMEAGPMVRKGTM 482

DB 417 PNTLKGIMDEERTISNLRSICIAQRNVAFINTGFLDRTGDEMHSWAGPMLRNQM 476

QY 483 KSQPMILAYEDPHNVDAAGLAFSGRAQVGKGMVTMTLMADMVETKIAQPRAGASTAWVP 542

DB 477 KSTPMKAYERNVNLGSLFCGLRKAQIGKGMAMPDMADMYSKQDQLRAGANTAWVP 536

QY 543 SPTAATLHALHYQDVAAVQOGLA-----GKRATTIEQLLTPLAKELAWADEPREEV 597

DB 537 SPTAATLHALHYQTNVQSVQANTAQTEFNAEFPEPLDDLLTIPVAENANWSAQEQCEL 596

QY 598 DNNQCSILGYVVRVWDGVGSKVPDIHDVVALMEDRATLRISQALLANLWRHGVITTSADV 657

DB 597 DNNVQILGYVVRVWEQIGGSKVPDIHNVALMEDRATLRISQOHLANLWRHGLITKEQV 656

QY 658 RASLERMAPLVDRQAGDVAYRMAPNPFDDSIAPLAAQELILSGAQCPNGYTBILHR-R 716

DB 657 QASLENMAKVYDQQAGDPAIRPMWAGNFANSCAFKAASDLIFLGVKQPNGYTBELLHAMR 716

QY 717 RRE 719

DB 717 LRE 719

RESULT 10

SI7774

malate synthase (EC 4.1.3.2) - *Neurospora crassa*

C:Species: *Neurospora crassa*

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-May-2000

C:Accession: SI7774

R:Sandeman, R.A.; Hynes, M.J.; Fincham, J.R.S.; Connerton, I.F.

Mol. Gen. Genet. 228, 445-452, 1991

A:Title: Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and

A:Reference number: SI7773; MUID:91375430; PMID:1832736

A:Accession: SI7774

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <SAN>

A:Cross-references: EMBL:X56672; NID:92974; PIDN:CAA39994.1; PID:92975

C:Genetics: 447/3

A:introns: 447/3

C:Superfamily: malate synthase

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 5.4%; Score 205.5; DB 2; Length 542;

Best Local Similarity 24.4%; Pred. No. 6.6e-06;

Matches 135; Conservative 64; Mismatches 205; Indels 149; Gaps 33;

QY 194 VODGOLVVALPKSTGLANPCQFAGYTGAAESPTSVLLINHGHIETILDPESQVGTDR 253

DB 58 IDRGVLPDFLPETKHIRENP---TWKGAA--PAAPPLVDRRVEV-----TGPTDR 102

QY 254 AGVKDVLIESAITTMDFEDSVAAVDAADKVLGYRNWLGKNGD-----LAAAVD-KDG 306

DB 103 KMNVALNSDVVTYNADFEDSSAP-----TWANMVNGQNLVDARQIDFKQG 151

QY 307 TAFRLVLRDRNTYAPGGGQFTLPGRSLMFVRNVG-HLMTNDATVDTDGSSEVFEGIMDAL 365

DB 152 PREYK-LATDR-----TLP--TLIVRPRGWL--EEKHTVIDGEVSGSLFD-- 193

QY 366 FTGLIAIHLKASDVNGPLINSRTGSIYIVKPKHGPFAVFTCELSRVEDVLGLPQNT 425

DB 194 -FGLYFFFNK-----ELVQGFQF-YFYPKPMESHLEARLWNAFNAQDYVGIPLST 245

QY 426 MKIGTMDERTTNNLKACIKAAADRVVFINTEGFLDRTG-DEIHTSMEAGPMV--- 477

DB 246 I-----RGTVLITETITAAFEWDEIIFELRNHTSLNGGWDYIF-----PFIKEV 290

QY 478 -----RKGTWKSQFWILAYED-----HNV-----DAGLAAGFSGRAQVGKGMVTMT 518

DB 291 RFPNVLPRSDVTMTVPFMEAYVKLLIKTLHLVWAMGAAQIPIKDDKAANDKAME 350

QY 519 ELKADWETKIAQPRAGASTAWPSPATAATL-----HALHYQDVAAVQOGL 566

DB 351 GVRAD-----KUREAPAGHDGTWVAHPALASIALEVFNKHMTPTNQLFNRREDVKIGQDL 406

QY 567 AGKRATTIEQLLTPLAKELAWADEPREEVNDCQSILGYVVRVWDGVGSKVPDIHD 626

DB 407 LN-----MNVPGSS-----TEDGIRK-----NLNTGLGYTEPMI-RGVGC--VPKHP 446

QY 627 VALMEDRATLRISQALLANLWRHGVITTS-----ADVRSASLERMAPLVDRQAGDVAYRPM 682

DB 447 ---QEDALTAIEVRSQQLWQVWVHRTTAEKHVYKRYPL-KLLKEADRQR---LAKAPQ 499

QY 683 PNFDDSIAPLAAQ 695

DB 500 NKFNLAAQVYFASQ 512

RESULT 11

D91245

malate synthase A [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D91245
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D91245
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <HAY>
A:CROSS-references: GB:BA000007; PIDN:BA38355.1; PID:gl3364408; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs4932
C:Superfamily: malate synthase

Query Match 5.0%; Score 190.5; DB 2; Length 533;
Best Local Similarity 22.8%; Pred. No. 6.4e-05;
Matches 117; Conservative 67; Mismatches 241; Indels 89; Gaps 21;

QY 159 YNKVRGDKVIYARKFLDDSVPLSSGSGFGDATGFTVQDQGVVLPDKSTGLANPGQFAG 218
DB 17 YGQEQKQILTAZAVFELTV-----THFTPQNKLLAARIQQQODIN-GTLPD 65

QY 219 YTGAAESPTSVLLINHLHIEILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAV 278
DB 66 FSETASIRDADWKIRGIPADLEDRVEITGSEVERKWINALNANVKVFMADFEEDSLAP- 124

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIR-VLNRDRNYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNEAGKIYQKPNPAVLI 164

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIR-VLNRDRNYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNEAGKIYQKPNPAVLI 164

QY 338 RNUGHLMTNDIVDTDGSEVFEGIMDALFT-GLIAIHGLKA--SDVNGPLINSRTGSIYI 394
DB 165 CRVRGLHLPKHWTRG-----EALPGSLFDPALYFFHNYQALLAKGSGP-----YF 211

QY 395 VKPQHGPAEVAFTCELFSEVEDVGLPONTMKIGIMDEERTTVNLKACIKAAADRVVF 454
DB 212 YLPKTSQWQEAAMWSEVFSYAEDRNLPRGTIKATLIELTLPVAFQMDIHALRDHVG 271

QY 455 INTGFLDRTGDEHTSMEAGPMV---RKGTMSQPMILAYED-----HNVDA-----GLA 501
DB 272 LNCGRWDYIFSYIKTLKNYPDRVLPDRQAVTMDKPFNLNAYSRLLIKTCHRGCAFAMGWA 331

QY 502 AGFSGRAQVKGKMTWELMADVETKIAQPRAGASTAWVPSPPTAATLHALHYQVDVA- 560
DB 332 AFIPSKDEEHNQ-VLNKVKAD---KSLAANNHGDGTWIAHPGLA-----DTAM 376

QY 561 AVQOGLAGKRATIEQL--LTIPLAKELAWAP---DEIREEDNNCSILGYVVRWVDQ 615
DB 377 AVFNDILGSRKNQLEVRREQDAPITADQLLAPCDGERTEGMRANIRVAVQVIEAWI-SG 435

QY 616 VGCSKVPDIHDVALMEDRATLRISQLLANLWRH 649
DB 436 NGCVPI-----YGLMEDAATAEISRTSIQWIIH 464

RESULT 12
B86093
malate synthase A [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B86093
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B86093
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-533 <STO>
A:CROSS-references: GB:AE005174; NID:gl2518944; PIDN:AAGS9206.1; GSPDB:GN00145; UWGP:Z56
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: aceB
C:Superfamily: malate synthase

Query Match 5.0%; Score 190.5; DB 2; Length 533;
Best Local Similarity 22.8%; Pred. No. 6.4e-05;
Matches 117; Conservative 67; Mismatches 241; Indels 89; Gaps 21;

QY 159 YNKVRGDKVIYARKFLDDSVPLSSGSGFGDATGFTVQDQGVVLPDKSTGLANPGQFAG 218
DB 17 YGQEQKQILTAZAVFELTV-----THFTPQNKLLAARIQQQODIN-GTLPD 65

QY 219 YTGAAESPTSVLLINHLHIEILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAAV 278
DB 66 FSETASIRDADWKIRGIPADLEDRVEITGSEVERKWINALNANVKVFMADFEEDSLAP- 124

QY 279 DAADKVLGYRNWGLNKGDLAAVADKGTAFIR-VLNRDRNYTAPGGQFTLPGRSLMFV 337
DB 125 -----DW---NK-----VIDGQINLRDAVNGTISYTNEAGKIYQKPNPAVLI 164

QY 338 RNUGHLMTNDIVDTDGSEVFEGIMDALFT-GLIAIHGLKA--SDVNGPLINSRTGSIYI 394
DB 165 CRVRGLHLPKHWTRG-----EALPGSLFDPALYFFHNYQALLAKGSGP-----YF 211

QY 395 VKPQHGPAEVAFTCELFSEVEDVGLPONTMKIGIMDEERTTVNLKACIKAAADRVVF 454
DB 212 YLPKTSQWQEAAMWSEVFSYAEDRNLPRGTIKATLIELTLPVAFQMDIHALRDHVG 271

QY 455 INTGFLDRTGDEHTSMEAGPMV---RKGTMSQPMILAYED-----HNVDA-----GLA 501
DB 272 LNCGRWDYIFSYIKTLKNYPDRVLPDRQAVTMDKPFNLNAYSRLLIKTCHRGCAFAMGWA 331

QY 502 AGFSGRAQVKGKMTWELMADVETKIAQPRAGASTAWVPSPPTAATLHALHYQVDVA- 560
DB 332 AFIPSKDEEHNQ-VLNKVKAD---KSLAANNHGDGTWIAHPGLA-----DTAM 376

QY 561 AVQOGLAGKRATIEQL--LTIPLAKELAWAP---DEIREEDNNCSILGYVVRWVDQ 615
DB 377 AVFNDILGSRKNQLEVRREQDAPITADQLLAPCDGERTEGMRANIRVAVQVIEAWI-SG 435

QY 616 VGCSKVPDIHDVALMEDRATLRISQLLANLWRH 649
DB 436 NGCVPI-----YGLMEDAATAEISRTSIQWIIH 464

RESULT 13
AD1011
malate synthase (EC 4.1.3.2) - Salmonella enterica subsp. enterica serovar Typhi (strain
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AD1011
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, P.; Davis, P.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD1011
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-533 <PAR>
A:CROSS-references: GB:ALU13382; PIDN:CAD09189.1; PID:gl6505193; GSPDB:GN00176
C:Genetics:
A:Gene: STY4401
C:Superfamily: malate synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 5.0%; Score 190; DB 2; Length 533;

Best Local Similarity 23.2%; Pred. No. 6.9e-05;
Matches 138; Conservative 67; Mismatches 249; Indels 140; Gaps 27;

QY 158 TYNKVRG-----KVIAYARKFLDSDVPLSSGSGFGATGTTVDGGLVVALPKSTGLAN- 212
DB 12 TFRPQGLEKQVLTAEEVSLTEL-----TFTPKRNKLLAARTCQOQDINDG 61
QY 213 --PGFAGYTGAAASPTSVLLINHGL---HIEILLDPESQVGTTRAGVKDVILSAITT 267
DB 62 KLPDFISETTGIRESNMQIRGIPADLQDRRVEI-----TCPVERKMVINALNAVVF 114
QY 268 IMDFEDSVAADAAADKVLGYRNWLGKNGDLAAAVDXDGTAFLR-VLNDRDNYTAPGGQ 326
DB 115 MADFEDSLAP-----DW---NK-----VIDQINLRADAVNGTISYINEAGKI 153
QY 327 FTLPGRLMVFYRNHGLMTNDAIVDTDGSEVFEGIMDALFT-GLIAIHGLKA---SDVNGP 383
DB 154 YQLKXPDPAVLICRVGRGLHLPKHKVTRWG---EAPGSLPDFALYFFHNYKALLAKGSGP 209
QY 384 LINSRTGSIYIVKQKMGCPAEVATCELSRVEDVLGLPONTMKIGIMDEERRITVLKA 443
DB 210 -----YYLPKTKQWQAANWSEVFSVAERFNLPRGTIKATLIETLPVAFQWDE 260
QY 444 CIKAAADRWWPINTGFLDRTGDEHTSMAGPMV---RKGTMKSQPMILAYED-----H 494
DB 261 ILHALRDHIVGLNCRWDYIFSYIKTLKNHPDRVLDPQVVVMDKPFLSAYSRLLIKTCH 320
QY 495 NVDA---GLAGFGSRAQVKGKMTMELMADNWETKIAOPRAGASTAWPSPAAITLHA 551
DB 321 KRGAFAFMGMAAFIPSKDVERNNQVLAKVKAD---KALEANNHGHDGTWTAHP----- 369
QY 552 LHYHQVDVAAVQQGLAGKRRTIETQLL-----TIPLAKELAWAP---DEIRE 595
DB 370 -----GLADTAMAVFNEVJGEHKNQLFITREDAPIAQLLEPCGEGETEA 416
QY 596 EVDNNCSILGYVVRWDQGVGCSKVPDHDHVALMEDRATLIRISSOLLANLHR----- 649
DB 417 GMRANIRVAVGYIEAWT-PGNGCVPI-----YGLMEDAATAEISRTSIGWIIHKTLSN 470
QY 650 -GVITTSADVRSAL-ERMAPLVD-----RONAG--DVAVRPM--APNFDDSDIAFL 692
DB 471 GKPYTKALFREMIAEEMVIODELGEHRYSSGRFDDARLMEQITTSDDLIDFL 524

RESULT 14
SYECMA
malate synthase (EC 4.1.3.2) A - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 31-Mar-1990 #sequence revision 31-Mar-1990 #text_change 01-Mar-2002
C:Accession: A32649; A30378; E65208; Q00592
R:Byrne, C.
submitted to the EMBL Data Library, July 1988
A:Reference number: S05692
A:Accession: A32649
A:Molecule type: DNA
A:Residues: 1-533 <BYR1>
R:Byrne, C.R.; Stokes, H.W.; Ward, K.A.
Nucleic Acids Res. 16, 10924, 1988
A:Title: Nucleotide sequence of the aceB gene encoding malate synthase A in Escherichia
Science 277, 1453-1462, 1997
A:Reference number: A30378; MUID:89083515; PMID:3060852
A:Accession: A30378
A:Molecule type: DNA
A:Residues: 1-533 <BYR2>
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65208
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-533 <BLAT>
A:Cross-references: GB:AE000474; GB:U00096; NID:gl790440; PIDN:AAC76984.1; PID:gl790444.1

A; Experimental source: strain K-12, substrain MG1655

C;Genetics:

A; Gene: aceB

A;Map position: 91 min

C;Superfamily: malate synthase

C; keywords: carbon-carbon lyase; coenzyme A; glyoxylate bypass; oxo-acid-lyase

Query Match

Best Local Similarity 22.7%; Pred. No. 7.5e-05;

Matches	120;	Conservative	67;	Mismatches	253;	Indels	89;	Gaps	21;
---------	------	--------------	-----	------------	------	--------	-----	------	-----

QY 144 TDVIPETDGAEGPTYNKKVRGDKVIAYARKFLDDSVPLSSGSEGDATGFTVQDGLVVAL 203

[illegible]

.....

QY 204 PDKSTGLANPGQFAGYTGAAESPTSVLLINHGLHIEILIDPESQVGTTRAGVKDVILES 263.

Db 52 IQQQQDIDN-GTLPDFISETASIRDADWKIRGIPADLEDRRVEITGPVERKVINAN 110

Q- 364 A T T T I M P E E S S I A A I N D A A D Y A I T C Y P A N T I C I N K A C H I A A I N D A A D Y A I T C Y P A N T I A D 333

111 WKVENBDEESLB-:||||:|
:||||:|
DW--NY--WINDCOTINI BBAVNSTCTSYTNE 118

DD . III VAVMAFEDSLAF-----DW---NK-----VIDGQINLRDAVNGTISYINE 149

QY 323 GGGQF1LFGRSLMFVRNVGHLWINDAIVDIDGSEVFEGINDALF1-GLTATHGLKA--SD 379

D5 AGKIYQLKPNPAVLICRVRLHLPEKHVTWRG----EAI PGLSFDFAFYFFHNYQALLAK 205

Qy 380 VNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLPQNTMKIGIMDEERRTV 439

[illegible]

QY 440 NLKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMV---RKGTMKSQPWILAYED--- 493

D_b 257 QMDEILHALRDHIVGLNCGRWYIFS^{YIK}TLKNYPDRVLPDRQAVTMDKPF^{LNAYSRL}LI 316

494 --HNBD2-- --CTACGECBPACVCKGMMTMTETMA DMVETVTAADBPACGCTATGDSPTA 545

317 КТСНКСВБЗМССМДЗБІРСКДБЕЕННО-УІ НКМ'УД- --- КСТ БАННСНДСТІАНРСІ. 371

DD 317 AICHKXGAFATIGOMTAF IFSDWEHNNQ=VDNNVND-----NSDEANNHGHDGIWIAHFGI 371

RESULT 15

H82287

malate synthase A VC0734 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C;Accession: H82287

R;Heidelberg, J.F.; Exton

1. R.R.: Mekalanos, J.J.

Nature 406, 477-483, 2001.

A;Title: DNA Sequence of

A;Reference number: A82

A;Accession: H82287

A;Status: preliminary
λ: Molecule type: DNA

A;Molecule type: DNA
A:Residues: 1-556 <HET>

A; Restraques: 1-390 <HBI>
A; Cross-references: GB: A

A;Experimental source: 5

C;Genetics:

A;Gene: VC0734

Search completed: November 21, 2003, 16:09:50
Job time : 18.4376 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copy-right (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 8.72562 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-2

Perfect score: 3810

Sequence: 1 TDRVSVGNLRVLYDFVN.....KARAAEKPAAPSRRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	100.0	741	1 MASZ MYCTU	Q50596 mycobacteri
2	3117.5	81.8	731	1 MASZ MYCLE	Q32913 mycobacteri
3	2634	69.1	724	1 MASZ RHOPA	Q9ae55 rhodococcus
4	2569	67.4	725	1 MASZ PSEAE	Q9i636 pseudomonas
5	2541	66.7	725	1 MASZ PSEPK	Q86cx8 pseudomonas
6	2497	65.5	725	1 MASZ PSEFL	Q05137 pseudomonas
7	2489	65.3	725	1 MASZ PSESM	Q88ab2 pseudomonas
8	2416	63.4	721	1 MASZ RHIL0	Q98d4 rhizobium l
9	2372.5	62.3	731	1 MASZ AGRT5	Q8u185 agrobacteri
10	2367.5	62.1	723	1 MASZ RHILV	Q937w7 rhizobium l
11	2351.5	61.7	723	1 MASZ RHIME	Q92ta4 rhizobium m
12	2318	60.8	728	1 MASZ BRUME	Q8yir3 brucella ne
13	2313	60.7	728	1 MASZ BRUSU	Q8fz50 brucella su
14	2289	60.1	738	1 MASZ CORGL	P42450 corynebacte
15	2273	59.7	748	1 MASZ COREF	Q8fnb3 corynebacte
16	2272.5	59.6	721	1 MASZ BRAJA	Q89ue3 bradyrhizob
17	2256.5	59.2	727	1 MASZ BACHD	Q9kb03 bacillus ha
18	2215	58.1	725	1 MASZ PSESM	Q87z72 pseudomonas
19	2168.5	56.9	722	1 MASZ ECOL6	Q8fdn6 escherichia
20	2156.5	56.6	722	1 MASZ SHIFL	P59663 shigella fi
21	2154.5	56.5	722	1 MASZ ECOLI	P73330 escherichia
22	205.5	5.4	542	1 MASZ NEUCR	P28345 neurospora
23	193.5	5.1	1005	1 GCP CAEEL	Q10663 caenorhabdi
24	189.5	5.0	533	1 MASZ ECOLI	P08997 escherichia
25	185	4.9	551	1 MASZ CANTR	Q02116 candida tro
26	184.5	4.8	541	1 MASZ MYXXA	P95329 myxococcus
27	175	4.6	528	1 MASZ EMENI	P28344 emericella
28	171.5	4.5	559	1 MASZ MAIZE	P49081 zea mays (m
29	171	4.5	554	1 MASZ YEAST	P21826 saccharomyc
30	165	4.3	554	1 MASZ YEAST	P30952 saccharomyc
31	160	4.2	555	1 MASV PICAL	P21360 pichia angu
32	153.5	4.0	541	1 MASV STRCL	Q9z477 streptomyce
33	151.5	4.0	564	1 MASV SOYBN	P45458 glycine max

ALIGNMENTS

RESULT 1

MASZ_MYCTU

ID MASZ_MYCTU STANDARD; PRT; 741 AA.

AC Q50596;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Malate synthase C (EC 2.3.3.9).

GN GLCB OR RV1837C OR M1885 OR MCV1A11.06.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RL complete genome sequence.";

RN Nature 393:537-544(1998).

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Ailand D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

RL laboratory strains.";

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +

CoA.

CC -!- PATHWAY: Glyoxylate bypass; second step.

CC -!- SUBUNIT: Monomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL; 278020; CAB01465.1; -.

DR EMBL; AE007047; AAK46156.1; -.

```

DR PIR; F70722; F70722.
DR PDB; IN81; 18-DEC-02.
DR PDB; IN8W; 18-DEC-02.
DR TIGR; MT1885; -.
DR Tuberculin; Rv1837c; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfam; TIGR01345; malate syn G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome; 3D-structure.
FT ACT_SITE 339 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 633 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 741 AA; 80403 MW; A32F54E0FE8B7C64 CRC64;

Query Match 100.0%; Score 3810; DB 1; Length 741;
Best Local Similarity 100.0%; Pred. No. 2 5e-241;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 60
Db 2 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPQLVVPV 120
Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPQLVVPV 121
QY 121 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDSDVP 180
Db 122 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDSDVP 181
QY 181 LSSSGFDGATGFTVDQGLVVALPKSTGLANPQFAGYGAESPVSLLINHLGHIEI 240
Db 182 LSSSGFDGATGFTVDQGLVVALPKSTGLANPQFAGYGAESPVSLLINHLGHIEI 241
QY 241 LIDPESQVGTTRDAGVXDVIIESAITTIMPDESVAAADKVLGVNMLGLNKGDLLA 300
Db 242 LIDPESQVGTTRDAGVXDVIIESAITTIMPDESVAAADKVLGVNMLGLNKGDLLA 301
QY 301 AVDXDGTAFRLVNRDNYTAPGGQQTLPGRSLFVRNVGHLMNTDAIVDTGSEVFEG 360
Db 302 AVDXDGTAFRLVNRDNYTAPGGQQTLPGRSLFVRNVGHLMNTDAIVDTGSEVFEG 361
QY 361 IMDALFTGLIAIHGLKASDVNGPLINSGTSGIYIVKPMHGPAEVAFTCELFSRVEDVLG 420
Db 362 IMDALFTGLIAIHGLKASDVNGPLINSGTSGIYIVKPMHGPAEVAFTCELFSRVEDVLG 421
QY 421 LPQNTMKIGIMDEERTTVNLKACIKAAADRVVPEINTGFLDRGTDEIHTSMEAGPMVRKG 480
Db 422 LPQNTMKIGIMDEERTTVNLKACIKAAADRVVPEINTGFLDRGTDEIHTSMEAGPMVRKG 481
QY 481 TMKSQPTLAVEDHNVDAGLAAGSGRAOVCKGMWTELMADWVETKIAQPRAGASTAW 540
Db 482 TMKSQPTLAVEDHNVDAGLAAGSGRAOVCKGMWTELMADWVETKIAQPRAGASTAW 541
QY 541 VPSPTAATLHALHYQVDVAQVQGLAGKRATIEQLTIPLAKELAWAPDEIREEDNN 600
Db 542 VPSPTAATLHALHYQVDVAQVQGLAGKRATIEQLTIPLAKELAWAPDEIREEDNN 601
QY 601 COSILGVVVRWDQVCGSKVPDTHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 660
Db 602 COSILGVVVRWDQVCGSKVPDTHDVALMEDRATLRISQLLANWLRHGVITSADVRAS 661
QY 661 LERVAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGCYTEPILHRRREF 720
Db 662 LERVAPLVDRQAGDVAYRPMAPNFDDSIAPLAAQELILSGAQPNGCYTEPILHRRREF 721
QY 721 KARAAEPAPSDRAGDDAAR 740
Db 722 KARAAEPAPSDRAGDDAAR 741

```

RESULT 2
 MASZ MYCLE
 ID MASZ MYCLE STANDARD; PRT; 731 AA.
 AC O32913.
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR ML2069 OR MLCB1788.27.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1789;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011 (2001).
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CoA.
 CC -!- PATHWAY: Glyoxylate bypass; second step.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the malate synthase family. GlocB subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL008609; CAAL5459.1; -.
 DR EMBL; AL583924; CAC31024.1; -.
 DR EIR; T44752; T44752.
 DR HSSP; P37330; 1D8C.
 DR Leptoma; ML2069; -.
 DR HAMAP; MF_00641; -; 1.
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR006253; Malate synthase.
 DR Pfam; PF01274; Malate synthase; 1.
 DR TIGRfam; TIGR01345; malate syn G; 1.
 KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
 KW Complete proteome.
 FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 638 638 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 731 AA; 80141 MW; 3878CADA5DB416C CRC64;

Query Match 81.8%; Score 3117.5; DB 1; Length 731;
 Best Local Similarity 82.0%; Pred. No. 4 3e-196;
 Matches 597; Conservative 54; Mismatches 77; Indels 5; Gaps 2;

QY 1 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 60
 Db 2 TDRVSGNLRIRARVLYDFVNNALPGTIDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
 QY 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPQLVVPV 120
 Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTAGPQLVVPV 121
 QY 121 LNARFALNAANARWGSGLYDALYGTDIVIPETDGAEGKGYTNKVRGDKVIAYARKFLDSDVP 180

Db 122 LNARFALNAANARWGSYLDALYGTDTIPETEGAEGSEYKIRGDKVIAVARKFMDQAVP 181
 QY 181 LSSGSEDAFCFTVQDQOLVVALPDKSTGLANPQAGTGAEST-SVLLNHGLHIE 239
 Db 182 LADSMTNAGVSIFDQGLQIATGNTSTGLASEKEFVGNRQLRSSNWSVLLANHGLHIE 241
 QY 240 ILIDPESQVGTDRAGVKDVIKLESATTTIMDFDSVAADKVLGYRNWGLNKGDLA 299
 Db 242 VLIDPESPIKTPVGKIVKIVLESATTTIMDFDSVTAVDADKVGYNWGLNKGDLT 301
 QY 300 AAYDKGTAFRLVNRDRNTAPGGQFTLPGRSLMFVRNVGHLMNDIAIV---DTDGS 355
 Db 302 ERYNKDCKTFTRVINDRSYTTDPGGELTLPGRSLLFVRNVGHLLTSDAILYDGGDQGBK 361
 QY 356 EVEGIMDALFTGLIAHGLKASDVNGPLNSITGSIYIVKPMHGPAAEVATCELFSSV 415
 Db 362 EVEGIIIDAVFTGLAIAHGLKGTGEANGPLNSITGSIYIVKPMHGPAAEVATCELFSSV 421
 QY 416 EDVLGIPQNTWKIGIMDEERTTNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGP 475
 Db 422 EDVLGLPQGTGLKVGIMDEERTTNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGP 481
 QY 476 MVKSGTKMKSOPWTLAYEDHNVDAGLAAGFSGRAQVKGWMTYTMELMADVETKIAQPRAG 535
 Db 482 MIRKGAKNSTWIKAYEDANVDIGLAAGFKGAQIGKGMWMTYTMELMADVETKIAQPRAG 541
 QY 536 ASTAWPSPATATLHALYHVDVAAVQOGLAGKREATTIEQLITIPLAKELAWAPDEIRE 595
 Db 542 ATTAWPSPATATLHAMHYHVDVAAVQOGLTQORRATVDQLITIPLAKELAWAPDEIRE 601
 QY 596 EVDNQCOSILGYVVRVWDQGVGSKVPDTHVVALMEDRATLRISSOLLANWLRHGVITSA 655
 Db 602 EVDNQCOSILGYVVRVWDQGVGSKVPDTHVVALMEDRATLRISSOLLANWLRHGVITSE 661
 QY 656 DVEASLERMAPLDVDRQVAGVAVRPMAPNFDDISIAFLAQELILSQAQPNGYTEPILHR 715
 Db 662 DVEASLERMAPLDVQVNAEDPARVPMAPNFDDISIAFLAQELILSQAQPNGYTEPILHR 721
 QY 716 RRREFKAR 723
 Db 722 RRREFRAQ 729

RESULT 3

MASZ_RHOFA STANDARD; PRT; 724 AA.
 ID MASZ_RHOFA
 AC Q9AES5;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR VICA.
 OS Rhodococcus fascians.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Nocardiaceae; Rhodococcus.
 OX NCBI_TaxID=1828;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D189;
 RA Verecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Hoisters M.,
 RA Goethals K.;
 RA "Characterization of a chromosomal locus that affects pathogenicity in
 RT Rhodococcus fascians";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CC CoA.
 CC -1- PATHWAY: Glyoxylate bypass; second step.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AJ301559; CAC35701.1; -
 DR HSSP; P37330; 1D8C;
 DR HAMAP; MF_00641; 1;
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR006253; Malate synthase.
 DR Pfam; PF01274; Malate synthase; 1.
 DR TIGRFAMs; TIGR01345; malate syn G; 1.
 KW transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
 FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 724 AA; 78609 MW; F889FE883890995E CRC64;

Query Match 69.1%; Score 2634; DB 1; Length 724;

Best Local Similarity 69.0%; Pred. No. 1.6e-164;

Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;

QY 1 TDRSVGNLRARLVLYDFVNNALPGTIDPDSTWAGVDKVVADLTTONQALLNARBELQ 60
 Db 2 TDRVQAGLQVAKVLFDFVEKEALPGTIDLDSEAFWAGAAASVIADLAPKXKALLAVRDEIQ 61
 QY 61 AQIDKWHRRRVIEPIDMDAYRQFLTETGYLLPEPDDFTITTSVGDABEITTTAGPOLVVPV 120
 Db 62 GKVDANHGEHAGAEVDRAAYKAFLEIGYLLDEADFQIHTSGVDTEITTTAGPOLVVPV 121
 QY 121 LNARFALNAANARWGSYLDALYGTDTIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVP 180
 Db 122 LNARFALNAANARWGSYLDALYGTDAIPETDGAEGKPTYNKRGDKVIAYARKFLDSDVP 181
 QY 181 LSSSFGDATGFTVQDQOLVVALPKST-GLANPQAGTGAEST-SVLLNHGLHIE 239
 Db 182 LSSSHVGTGTVVDAASLTVTLAGDSTVGLKDSQLLYGQGTDPATLTFVNHGLHIE 241
 QY 240 ILIDPESQVGTDRAGVKDVIKLESATTTIMDFDSVAADKVLGYRNWGLNKGDLA 299
 Db 242 ILIDPESPIKTPVGKIVKIVLESATTTIMDFDSVAADKVLGYRNWGLNKGDLT 301
 QY 300 AAYDKGTAFRLVNRDRNTAPGGQFTLPGRSLMFVRNVGHLMNDIAIVDTGSEVFE 359
 Db 302 BEVSKGKGTFTFRANKKORTYTSVDGSELTLHGRSLLFVRNVGHLMTSDAILDADGNEVPE 361
 QY 360 GIMDALFTGLIAHGLKASDVNGPLNSITGSIYIVKPMHGPAAEVATCELFSSVDEVL 419
 Db 362 GILDALFTSLAGLHSLTFDNLV---LSNSRTGSLYIVKPMHGPDEVAFTALFGRVEQVL 418
 QY 420 GLPQNTKIGIMDEERTTNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRK 479
 Db 419 GLPTNTLKVGIMDEERTTNLKACIQAASERVVFINTGFLDRTGDEIHTSMEAGPMVRK 478
 QY 480 GTMKSQPILAYEDHNVDAGLAAGFSGRAQVKGWMTYTMELMADVETKIAQPRAGASTA 539
 Db 479 GAMKGEKWIAYEDFNVDTLGAAQLGKGAQIGKGMWAMPDLNMDLMLEOKIGHPKAGANTA 538
 QY 540 WVPSTAAATLHALYHVDVAAVQOGLAGKREATTIEQLITIPLAKELAWAPDEIREEDVN 599
 Db 539 WVPSTAAATLHALYHVKVDVFAHQETAKAKRATVDEILEIPLAPSTDWTDEEKQNELDN 598
 QY 600 NCQSTLGVVVRVWDQGVGSKVPDTHVVALMEDRATLRISSOLLANWLRHGVITSAVRA 659
 Db 599 NSQSTLGVVVRVWDHGVGSKVPDINDIALMEDRATLRISSQFIANMRHGVITSEEQVRE 658
 QY 660 SLERMAPLDVDRQVAGVAVRPMAPNFDDISIAFLAQELILSQAQPNGYTEPILHRRRE 719
 Db 659 SLKRMAPVVDQRNASDPYKELAFDFTDNTIAFQAASDLIFQGTSPQPNGYTEPILHRRRE 718
 QY 720 FKA 722
 Db 719 YKA 721

```

RESULT 4
NASZ_PSEAE STANDARD; PRT; 725 AA.
AC Q91636;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR PA0482.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.D., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Gollery L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. Glcb subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE004485; AAG03871.1; -
CC PIR; H83586; H83586.
CC HSP; P17330; 1D8C.
CC HAMAP; MF 00641; -; 1.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthase.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_synG; 1.
CC Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 240 340 CATALYTIC BASE (BY SIMILARITY)
FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;

Query Match
Best Local Similarity 67.4%; Score 2569; DB 1; Length 725;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

QY 1 TDRVSGNLRARVLYDFVNNELPGTIDPDSFWAGYDKVADLTQNALNARELQ 60
DB 2 TERVQGLQVAKVLFDFVNNELPGTGVSDTFTWGAEVINDLPKAKALLAKRELQ 61
QY 61 AQIDKWHRRRVTEPIDMDAYQFTEIGYLLPEPDDFTITTSVGDVDAITTTAGPQLVVPV 120
DB 62 AKIDGWHQARAGQAHDAVYAKFLAEIGYLLPEAEDFQAGTQNVDDDEIARWAGPQLVVPV 121
QY 121 LNARFALNANRWGSLVDALYIGTDVIVPETGAEGKPTYNKVRGDKVIYARKELDDSV 180
DB 122 MNARFALNANRWGSLVDALYIGTDVIVSEEGGAEGKGYNVRGDKVIYAFARFALDEA 181

```

```

QY 181 LSSGSFGDATGETTVQDGGQIVVALPDKS--TGLANPQFAGYTGAAEPTSVLLINLGLHIE 239
DB 182 LESGSHVDATSYSVKNGALVVALKNGSETGLKNGAQFLAQDGAAPQAVLLKNGLHFE 241
QY 240 ILIDPESOVGTTDRAGVQKDVILESAITTIMDFESVAAVDAADKVLGVNRLWGLKNGDLA 299
DB 242 IQIDPSSPVGQTDAAAGVQKDVLEMEALTIMDCESSVAADADDKVIYRNWGLKNGDLA 301
QY 300 AAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVVGHLMNTDAIVDTGSEVFE 359
DB 302 EEVSKGSGSTFTRTMNPDVYTRADGSELTLGKRSLLFVRNVGHLMNTDAILDKGNVFE 361
QY 360 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSIYIVKPKMHGPAEVAFTCELFSRVED 417
DB 362 GIQDGLFTSLIAH-----DLNGNSTRKNSKNSGSIYIVKPKMHGPEEAFTNELFGRVED 416
QY 418 VLGLPQNTMKIGIMDEERTTVNLKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMV 477
DB 417 VLGLPRLNTLKVIGIMDEERTTVNLKACIKAAKDRVVFINTGFLDRTGDEIHTSMEAGAV 476
QY 478 RKGTKMSQPWILAYEDHNVDAAGLAAGFSGRAQVGKGMWMTLMADVYETKIAOPRAGAS 537
DB 477 RKGAKMSEKWI GAYENNVDVGLATGLQKQAIQIGKMWAMFDLMAAMLEQKIGHPLAGAN 536
QY 538 TAWFSPPTAATLHALHYHQVDVAAVQOQGLAGKRRATIEQLLTIPLAKELAWAPDEIREV 597
DB 537 TAWFSPPTAATLHALHYHKVDVFASQAEKLTASVDDILTIPAPNTNTWAEIKNEV 596
QY 598 DNNCSILGYVVRVVDQGVGSKVPDIHDVLMEDRATLRISSQLLANLWLGHVITSADV 657
DB 597 DNNAGILGYVVRVVDQGVGSKVPDIHDVLMEDRATLRISSQLLANLWLGHVITSADV 656
QY 658 RASLERMAPLVDRQNAVGVAYRPMAPNFDSDISAFIAQELILSGAQOQNGVTEPIHRRR 717
DB 657 VESLRKMAVVDRQNAVGSFSPRPAPNFDSDVAFQAALELVVEGTRQNGVTEPIHRRR 716
QY 718 REFRAK 723
DB 717 REFRAK 722

RESULT 5
NASZ_PSEPK STANDARD; PRT; 725 AA.
AC Q88QX8;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR PP0356.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Nceatl D., Wedler H.,
RA Lauber J., Stjepandic D., Hohnsels J., Strasz M., Heim S.,
RA Kiewitz C., Eisen J., Tammis K.N., Dueterhoeft A., Thummier B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).

```

```
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB016775; AAG5987.1; -
CC DR TIGR; PP0356; -
CC DR HAVAP; MF_00641; -; 1.
CC KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
CC Complete proteome.
CC FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 725 AA; 78346 MW; 8363F218E6116AE1 CRC64;
CC
CC Query Match 66.7%; Score 2541; DB 1; Length 725;
CC Best Local Similarity 67.1%; Pred. No. 1.9e-158;
CC Matches 486; Conservative 95; Mismatches 139; Indels 4; Gaps 2;
CC
CC QY 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTPOQALLNARDLQ 60
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 2 TGVYQVGLQVAKVLYDFVNEALPGTGIIVAEQFWAGAEKIINDLAPKNAKLRDELQ 61
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 61 AQIDKWHRRVIEPIDMDAYRQELTEGYLLPEPDDFTITTSQVDAEITTTAGPOLVVPV 120
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 62 AKIDAWQARKQAHDAAYAKFLOEIGYLLPQADDFQATQNVDBEIAHAGPOLVVPV 121
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 121 LNARFALNAANRWGSLYDALYGTDTVPETDGAEGKFTYKVRGDKVIAYARFLDSDVP 180
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 122 MNARFALNAANRWGSLYDALYGTDAISDEGGAEGQGYKVRGDKVIAFAFLDEAAP 181
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 181 LSSGSFGDATGFTVDQGLVVALPKDS-TGLANPGQFAGYTGAEPTSVLLINHGLHIE 239
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 182 LAGSHVDSVTGYRIEGGKLVVALKGGSTNGLRDAQLIGFHDGAAAPTAVLLKHNLHFE 241
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 240 ILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADKVKLYGRNWLGNKGLDA 299
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 242 IQVDASTPVGSTDAAGVKDILMESALTINDCEDSVAADVADDKVIYRNLGLMKGLDA 301
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 300 AAYDKQGTAFRLVNRDRNTAGGGQFTLPGRSLMFVRNVGHLMNTDAIVTDGSEVFE 359
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 302 ESVSKGKGTFTRTMNPDRYAAPNGGSVTLHGRSLLFVRNVGHLMNTNPAILDAAQGNIEPE 361
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 360 GIMDALFTGLIAJHGLKASDVNGPLNSRTGSTYIVKPKMGHGAFAVTCELPSRVEDYL 419
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 362 GIQDGLFTNLIAHLNGLNG---NTSRKNIRSGSVIIVKPKMGHGPVEVAFATEIPSQVEDLL 418
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 420 GLPONTWKIGIMDEERTTNLXKACIKAAADRVRVINTGFLDRTGDEIHTSMEAGPMVRK 479
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 419 GMPRNTVKVIMDEERTTNLXKCIKAAAEVVRVINTGFLDRTGDEIHTSMEAGAVVRK 478
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 480 GTWKSQPIILAYEDHNDVAGLAFSGRAQVQKGMVMTLMDVETKIAQPRAGASTA 539
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 479 GAWNKSEKIGAYENNNDVGLATGLOCRQAIQKGMWMPDLMAAMLEQKIAHPLAGANTA 538
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 540 WVPSPATLHALHYHQVDYAAVQOGLAGKRRATIEQLLTIPLAKELAWAPEIEEVDN 599
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 539 WVPSPATLHALHYHKVDVQARQRELASFTPASVDLILAIPLAATNWSAEIRNELDN 598
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 600 NCOSILGYVVRVDQGVGSKVPDIHDVLMEDRATLRISSQLLANWLHGHVITSADVRA 659
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 599 NAQGIILGYVVRVIDQGVGSKVPDIINNVLGIMEDRATLRISSQLLANWLHGHVITSADVRA 658
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 660 SLERMAPLVDRONAGVAVRPMAPNFDSDISAFILAAQELILSGAQOQNGYTEPILHRRREE 719
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 659 SLKEMAVVDDQVAGDPLRPMAPNFDNDVAFQAAVELVVEGKQKQNGYTEPVLHRRREE 718
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 720 FKAR 723
```

```
Db 719 FKAR 722
||||
RESULT 6
ID MASZ_PSEFL STANDARD; PRT; 725 AA.
AC 005137;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR FC2.4.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L6.5;
RA Artiguenave F.M., Delecu M., Vilagines R., Danglot C.;
RT "A functional glyoxylate bypass is mandatory for utilization of
RT alkanes by Pseudomonas fluorescens.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y11998; CAA72726.1; -
CC DR HSP; P37330; 1D8C.
CC DR HAMAP; MF_00641; -; 1.
CC DR InterPro; IPR001465; Malate synthase.
CC DR InterPro; IPR006253; Malate synthase.
CC Pfam; PF01274; Malate synthase; 1.
CC TIGRFAMs; TIGR01345; Malate syn G; 1.
CC KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle.
CC FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 725 AA; 78861 MW; 0C3326FE67C9B381 CRC64;
CC
CC Query Match 65.5%; Score 2497; DB 1; Length 725;
CC Best Local Similarity 66.7%; Pred. No. 1.5e-155;
CC Matches 483; Conservative 85; Mismatches 152; Indels 4; Gaps 2;
CC
CC QY 1 TDRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTPOQALLNARDLQ 60
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 2 TEHVQVGLQVAKVLYDFVNEALPGTGIIVAEQFWAGAEKIINDLAPKNAKLRDELQ 61
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 61 AQIDKWHRRVIEPIDMDAYRQELTEGYLLPEPDDFTITTSQVDAEITTTAGPOLVVPV 120
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 62 ARIDTWHTHAGQAHDPVAYKAFLODQIGYLLPEADFOASTQNVDBEIAHAGPOLVVPV 121
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 121 LNARFALNAANRWGSLYDALYGTDTVPETDGAEGKFTYKVRGDKVIAYARFLDSDVP 180
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 122 MNARFALNAANRWGSLYDALYGTDAISEADGAEGKGYKVRGDKVIAFAFLDEAAP 181
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 181 LSSGSFGDATGFTVDQGLVVALPKDS-TGLANPGQFAGYTGAEPTSVLLINHGLHIE 239
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 182 LAGSHVDSVTGYRIEGGKLVVALKGGSTNGLRDAQLIGFHDGAAAPTAVLLKHNLHFE 241
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC QY 240 ILIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADKVKLYGRNWLGNKGLDA 299
CC Db : : : : : : : : : : : : : : : : : : : : : : : : : :
CC
```


DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MLR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CX Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303059;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GICB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF003004; BAB51267.1; -.
DR HAMAP; MF 00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMS; TIGR01345; malate_syn.G; 1.
DR Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
KW ACT_SITE 338 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 721 AA; 78058 MW; 65376311A1E1BDF CRC64;

Query Match 63.4%; Score 2416; DB 1; Length 721;
Best Local Similarity 64.4%; Pred. No. 2.9e-150;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;

QY 1 TDRVSGNLRIRVLYDFVNNELPGTDIDPDSFWAGVGVKVDLTTPQCALINARDELQ 60
Db 2 TDRIEAGURIAELHDFVAGEALPGTGAADAFWSGSAIVHDLAPKRNALLKKRDAMQ 61

QY 61 AQIDKWHRRVIEPIDMDAYRQELTEIGYLLPPDDFTTTSGVDAEITTTAGPOLVVPV 120
Db 62 ERLDGWVRNG-APDMEVYKSLKEIGLVNPEGPAPFVSTDNVDEIAVAGPOLVVPV 120

QY 121 LNAFPALNANRWGLYDALYGTVDVTPEDGAEKGTNNKVRGDKVIAYAKFLDDSV 180
Db 121 MNARYALNANRWGLYDALYGTDAIPETGAEKGFENPARGAKVIAWAKDFLDQSV 180

QY 181 LSSGSPGDATGFTVODQQLVVALPDKSTGLANPGOPAGYTGAEPSVLLINHLGHI 240
Db 181 LTSGKAGVNLGSVANGALKLGAGAGTTLADPRQAGTGRGDAAPDVALVYKGLHIEI 240

QY 241 LIDPESQVGTDRAGVKDVLISAITIMDFEDSVAADAAQKVLGYRNWLGKNGDLAA 300
Db 241 VDRNNQIGRTDPAGTADVILESAITIQDCEDSVAADAAQKVVVYRNWLGKNGDLAE 300

QY 301 AVDKDGTAFILVLRNDRNTAPGGQFTLPGSLMFVNVGHLMINDAIVDTGSEVPEG 360
Db 301 EISKAGRSVRKLNADRAYTAPAGGOITVFGSLMLVRNVGHLMINPAILDGRGNEVPEG 360

QY 361 INDALFTGLIAIHGLKASDV--NGPLNSRTGSIYIVKPKHGPDAEVAFTCELSRVEDV 418
Db 361 INDAALTAIALH-----DVGPEGRANRANRAGSMYVVKPKHGPDAEVAFAVEIFDRVEAL 415

QY 419 LGLPONTMKIGIMDEERTTNLAKACIAAADRVVVFINTGFLDRTGDEIHTSMEAGPMVR 478
Db 416 LGMFNTIKMGIMDEERTTNLKEAIPAERVRVVFINTGFLDRTGDEIHTSMEAGPMIR 475

QY 479 KGTMSQPWILAYEDHNVNDAAGLAFSGRAQVGKGMWMTMELMADMVETKIAQPRAGAST 538
Db 476 KGMKQAAWISAYEAWNVDTGLECGLAGHAIQIGKGMWAMPDLMAAMLEQKIAHPKAGANT 535

QY 539 AWPSPPTAATHALHYHOVDVAQVQGLAGERRATIEQLLTPIPLAKELAWAPDIREVD 598
Db 536 AWPSPPTAATHALHYHVDVHVAQAALKSRPKAKLDDILSVFVAVRPNWTPDEIQRELD 595

QY 599 NNCOSILGVYVRVDQGVGCKVPDIHDVLMEDRATLRISSQLLANWLRHGVITSADVR 658
Db 596 NNAQILGVYVRWIDQGVGCKVPDINDVGLMEDRATLRISSQHLANWLRHKVCSEIQVR 655

QY 659 ASLERMAPLVDRONAGDVAYRPMAPNPDSDTAFLAAQELILSGACQPNGYTEPILHRRR 718
Db 656 DSLQMAAIVDRONVGDPLRYRMAPDFDKSIATQACDLVFKGTSQPNGYTEPVLHARRL 715

QY 719 EFKAR 723
Db 716 ELKAQ 720

RESULT 9
MASZ AGRTS
ID MASZ AGRTS STANDARD; PRT; 731 AA.
AC Q8U085;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
OS GICB OR ATU0047 OR AGR_C 78.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Marks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.,
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.

```

CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity). GlcB subfamily.
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008979; AAL41078.1; ALT INIT.
DR EMBL; AE007947; AAK85871.1; ALT_INIT.
DR PIR; AH2582; AH2582.
DR PIR; F97364; F97364.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthG.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfam; TIGR01345; Malate syn.G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT SITE 346 CATALYTIC BASE (BY SIMILARITY).
FT ACT SITE 637 CATALYTIC ACID (BY SIMILARITY).
FT ACT SITE 637 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 731 AA; 79520 MW; 69F304D5D6F8EFB CRC64;
Query Match 62.3%; Score 2372.5; DB 1; Length 731;
Best Local Similarity 63.6%; Pred. No. 2e-147;
Matches 461; Conservative 92; Mismatches 165; Indels 7; Gaps 3;
QY 1 TDRVSGNLRIRLVYDFVWNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDLQ 60
DB 13 TDKP---GLSIDRLIAYFLTDLVPLGTGLDSETFEGFSAIVHLSLSPKRELLAKRDALQ 69
QY 61 AQIDKWHRRVIEIDMDAYQFITEIGYLLPEPDDTTITTSVDAITTAGPOLVVPV 120
DB 70 EKIDGWYENG-APSDPDYAEFLKEIGYLLPEPGFKVETNNVDPEIAVVGQQLVVPV 128
QY 121 LNARFALNAANRWGSLYDALYGTVDVTPETDCAKGYPTYNKVRGDKVIARFELDDSV 180
DB 129 MWARFALNAANRWGSLYDALYGTDAISDAGAKGKGYNPKGDKVIARFELDESAP 188
QY 181 LSSGFGDATGFTVDDGQLVVALPDKSTGLANPQFAGYTGAAKPSPTSVLLINHLHIEI 240
DB 189 LETGSWSDVTGFTNADGGLQALGAATTLGLKDAVQFKGFGSEAAKPAITLLGNKGLHTEI 248
QY 241 LIDPESVGTTRDAGVKDVIKESATITIMDFEDSVAADAAKVLGYRNMLGNKGLDAA 300
DB 249 VIDPSTELGKDRAGISVILESATITIMDCEDSVAADAEKVLVYGNWMLGMRGLDTE 308
QY 301 AVDKDGTAFRLVLRNDRNYTPGGQFTLPGRSLMVRNVGHLMTNDAIVDTGSEVFEG 360
DB 309 AVSKGNTFTRLNDRYVYAPDGSALTLPGRSLMVRNVGHLMTNPAILDGRDVP 368
QY 361 IMDALFTGLIAHGLKASDVNGPLNSRTGSIVYKPKHGPAEVAFTCELSFSEVDVLG 420
DB 369 IMDAVVTALIALYDVGPS---GRRONSAGSMVYVVKPKHGPEEVAFAVELFVENVLG 425
QY 421 LPQNTWKIGIMDEERTTNNKACIAAADRVRVFINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 426 MAPNTMKGIMDEERTTNNKESIRAAKRVVFINTGFLDRTGDEIHTSMEAGPMVRKG 485
QY 481 TKKSQPVILAYEDHNVDAGLAAGFSGRAQVKGKMTWTELADMDVEIKIAPRAGASTAW 540
DB 486 DMKQAAWIAAYENWNVDTIGLECGLSGHAQIGKGMWAMPDLMAAMLEQKIAHPKAGANTAW 545
QY 541 VPSPTAATLHALYHQVDVAAVQOQGLAGKRATIBQLLTIPLAKELAWAPDEIEEVDNN 600
DB 546 VPSPTAATLHATHYHKVDVAAVQOGLKSGRAKLSILSVFVAPPNWTPPEIQLELDNN 605
QY 601 COSILGYVVRWVDQGVGCKVPDIDHVALMEDRATLRISQAQMANLRLHGHVTEAQI 660

```

```

DB 606 AQGILGYVVRWVDQGVGCKVPDINNIGLMEDRATLRISQAQMANLRLHGHVTEAQI 665
QY 661 LERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELTSCAQOPNGVTEPILHRRRREF 720
DB 666 MKRMAAVYDTQAGDPAYLPWASDFDGSVAFQAAVELVLKRGQPNGTPEVLRRLLEL 725
QY 721 KARAA 725
DB 726 KAKQA 730
RESULT 10
MASZ RHILV STANDARD; PRT; 723 AA.
AC Q937M7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=VF39;
RC Garcia de los Santos A., Hynes M.F.;
RT "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBSJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HAMAP; AY059637; AAL17965.1; -.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthG.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfam; TIGR01345; Malate syn.G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
FT ACT SITE 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT SITE 629 CATALYTIC ACID (BY SIMILARITY).
FT ACT SITE 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD6444 CRC64;
Query Match 62.1%; Score 2367.5; DB 1; Length 723;
Best Local Similarity 64.1%; Pred. No. 4.2e-147;
Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;
QY 3 RVSGNLRIRLVYDFVWNEALPGTDIDPDSFWAGVKVADLTQONQALLNARDLQ 62
DB 3 RVDKXGLAIETVLHDFLVEEVLPLGLAVDAKDFADFAIVHDLAPKNCALLAKRDQLVK 62
QY 63 IDKWHRRVIEPIDMDAYQFITEIGYLLPEPDDTTITTSVDAITTAGPOLVVPV 122
DB 63 IDWYRRHG-APADMDYOSFLREIGYLLPEGSDQVSTQNVDPFIAGIAGQLVVPV 121
QY 123 ARFALNAANRWGSLYDALYGTVDVTPETDCAKGYPTYNKVRGDKVIARFELDDSV 182
DB 122 ARYALNAANRWGSLYDALYGTDAIPESDCAEKSKYNPKRGEKVIARVRFELDTSAP 181
QY 183 SGSGFGDATGFTVDDGQLVVALPDKSTGLANPQFAGYTGAAKPSPTSVLLINHLHIE 241

```

Db 182 DCRWKDVGSAVKDGLVRSIDGQAMLTGDKHFGYRGDAAPATHILLKNGIHIEIV 241
Qy 242 IDPESVGTGTTDRAGVKDVLIESAITTIMDFEDSVAAVDAADKVLGVNWLGNKGLAA 301
Db 242 IDAATTIGKADSAHISDVLESIAITTIMDCEDSIAAVDAEDKVVVRNWLGNKGLQEE 301
Qy 302 VDKDGTAFRLVNRDRNTYAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSVEFEG 361
Db 302 VAKGTSFIRTLNPDLOYAGPDGAFAEVHRRSLMLVRNVGHLMTNPAILLDRDQNEVPEGI 361
Qy 362 MDALFTGLIALHGLKASDVNGPLNSHTGSIYIVKPMHGPAPVAFVTCFLSFVSDVGL 421
Db 362 MDAAITGLIALYDLPB---GRKNSRTGSMYVVKPMHGPPEVAFVAFVFSRVEDALGL 418
Qy 422 PONTMKGIMDEERRITVNLKACIAAADRWVINTGFLDRTGDEIHTSMEAGPMVRKGT 481
Db 419 PRNTIKMGIMDEERRITVNLKACIRARERVVINTGFLDRTGDEIHTSMEAGPMVRKGD 478
Qy 482 MKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADVETKIAOPRAGASTAW 541
Db 479 MRQAATISAYENWNVDIGLEGLAGHAQIGKGMWAMPDLMAAMLEOKIHPKAGANTAW 538
Qy 542 PPSAATLHALYHVQVDVAQVQGLAGKRRATIEQLTTLPLAKELAWAPDEIRVEVDNCC 601
Db 539 PPSAATLHALYHVVRNVNARVQGLKDRARAKLSILSVFVAVRPNWTPEEIQRELDNNA 598
Qy 602 QSLGIVYVRVWDQGVGSKVPDIHDVVALMEDRATLRISSOLLANWLHRGHVITSADVRASL 661
Db 599 QGILGYVVRVWDQGVGSKVPDINNVLGMDRATLRISSACHMANWLHKKVWTEAQIETW 658
Qy 662 ERMAPLVDRONAGDVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILHRRRREF 721
Db 659 RRMVAVDRONAGSDPAYRPMAGNFDSDIAFQAALDLVLKRGEPNGYTPFVLHRRRLEL 718
Qy 722 ARAA 725
Db 719 AKQA 722

RESULT 11
MASZ_RHIME STANDARD; PRY 723 AA.
ID MASZ_RHIME
AC Q92TA4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GUCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RC MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Fuenler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -|- PATHWAY: Glyoxylate bypass; second step.
CC -|- SUBUNIT: Monomer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the malate synthase family. GUCB subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL591782; CAC41449.1; -
CC HAMAP; MF 00641; -; 1.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthase.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate syn G; 1.
CC Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 338 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 78853 MW; A0E95B8A5164B58 CRC64;

Query Match 61.7%; Score 2351.5; DB 1; Length 723;
Best Local Similarity 62.9%; Pred. No. 4.6e-146;
Matches 455; Conservative 94; Mismatches 169; Indels 5; Gaps 3;

Qy 2 DEVSNGNLRIRARLVDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELOA 61
Db 2 DRVEKYQIDAGLHFRFLVEEAMFGTGVDAADRFFSFDLVHDLGPKNRALLVXRDELOA 61
Qy 62 QIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTTTSGVDATETTTAGPOLVWVL 121
Db 62 RLDGWYREH-APVDMEAYEAFLEIGYLLPEGDPFYVSTANVSEATATAGPOLVWV 120
Qy 122 NARFALNANARWGLYDALYDVTPIPTDGAEGKPTNKVGRDKVIAYAKFLDDSVPL 181
Db 121 NARYALNANARWGLYDALYDGTDAIETDGAERKGNPKRGAKVIAWAREFLDASAPL 180
Qy 182 SSGSGDGTATGTVDDGQLVVALPD-KSTGLANPQFAGYTCGAASPTSVLLINLGLHIEI 240
Db 181 AAGRWSDAKSFVEGATLVTADTKSAPRNSVQFAGYAGDPAAPSEIVLRNGLHIVI 240
Qy 241 LIDPESVGTGTTDRAGVKDVLIESAITTIMDFEDSVAAVDAADKVLGVNWLGNKGLAA 300
Db 241 VLDATTPIGKADAAGISDVLESIAITTIMDCEDSIAAVDAEDKVLVVRNWLGNKGLDLE 300
Qy 301 AVDKDGTAFRLVNRDRNTYAPGGQFTLPGRSLMFVRNVGHLMTNDAIVDTDGSVEFEG 360
Db 301 EVTKGGAFTRLNPDRAITAPDGAATLTPGRSLMVVRNVGHLMTNPAVLLDRDQNEVPEG 360
Qy 361 IMDALFTGLIALHGLKASDVNGPLNSHTGSIYIVKPMHGPAPVAFVTCFLSFVSDVGL 420
Db 361 LMDAMVTLIALHDIGR---NGRRANSRSGSMYVVKPMHGPPEVAFVAFVFAEALG 417
Qy 421 LPQNTMKGIMDEERRITVNLKACIAAADRWVINTGFLDRTGDEIHTSMEAGPMVRK 480
Db 418 LPANMKMGIMDEERRITVNLKACIRARERVVINTGFLDRTGDEIHTSMEAGPMVRK 477
Qy 481 TMKSQPWILAYEDHNVDAGLAAGSGRAQVKGKGMWMTLMADVETKIAOPRAGASTAW 540
Db 478 DMKQAPWISAYENWNVDIGLEGLAGHAQIGKGMWAMPDLMAAMLEOKIHPKAGANTAW 537
Qy 541 VPSAATLHALYHVQVDVAQVQGLAGKRRATIEQLTTLPLAKELAWAPDEIRVEVDN 600
Db 538 VPSAATLHALYHVVRNVNARVQGLKDRARAKLSILSVFVAVRPNWTPEEIQRELDNN 597
Qy 601 QOSILGYVVRVWDQGVGSKVPDIHDVVALMEDRATLRISSOLLANWLHRGHVITSADVRAS 660
Db 598 AQGILGYVVRVWDQGVGSKVPDINNVLGMDRATLRISSACHMANWLHGHVITSADVRAS 657
Qy 661 LERMAPLVDRONAGDVAYRPMAPNFDSDIAFLAAGELILSGAOPNGYTPFILHRRRREF 720
Db 658 MRMAAVDRONAGSDPAYRPMAGNFDSDIAFQAALDLVLKRGEPNGYTPFVLHRRRLEL 717
Qy 721 KAR 723
Db 722 KAR 723

718 KAK 720

RESULT 12

ID	MASZ	BRUME	STANDARD;	PRT;	728 AA.
DB	Q8YIR3;				
AC	15-SEP-2003 (Rel. 42, Created)				
DT	15-SEP-2003 (Rel. 42, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Malate synthase G (EC 2.3.3.9).				
GN	GLCB OR EME10380.				
OS	Brucella melitensis.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OC	Brucellaceae; Brucella.				
OX	NCBI_TaxID=29459;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=16M / ATCC 23456 / Biotype 1;				
RC	MEDLINE=20020109; PubMed=11756688;				
RA	DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Muij C., Los T., Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G., Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E., Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J., Haselkorn R., Kyripides N., Overbeek R.;				
RA	"The genome sequence of the facultative intracellular pathogen Brucella melitensis."				
RT	Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).				
RL	-!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate + CoA.				
CC	-!- PATHWAY: Glyoxylate bypass; second step.				
CC	-!- SUBUNIT: Monomer (By similarity).				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-!- SIMILARITY: Belongs to the malate synthase family. GCB subfamily.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).				
CC	EMBL; AE009480; AAL51561.1; -				
DR	PIR; AF3298; AF3299.				
DR	HMAP; MF_00641; -; 1.				
DR	InterPro; IPR001465; Malate synthase.				
DR	InterPro; IPR006253; Malate synthase.				
DR	Pfam; PF01274; Malate synthase; 1.				
DR	TIGRFAMs; TIGR01345; malate syn G; 1.				
KW	Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;				
KW	Complete proteome.				
FT	ACT_SITE 345 345 CATALYTIC BASE (BY SIMILARITY).				
FT	ACT_SITE 636 636 CATALYTIC ACID (BY SIMILARITY).				
FT	ACT_SITE 728 AA; 79985 MW; E1313A617979270D CRC64;				
SQ	SEQUENCE 728 AA; 79985 MW; E1313A617979270D CRC64;				
Query Match	60.8%; Score 2318; DB 1; Length 728;				
Best Local Similarity	61.8%; Pred. No. 7,2e-144;				
Matches	444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;				
QY	4 VSGNLRARVLYDFWNEALPGTDIDPDSFWAGVGVADLTQNOALLNARDELQAI 63				
DB	10 VIEIGLAVAPELVEFLAKEAPGCTGVEPEKFWKGFAMIRDLAPKNRALLAKDELQARI 69				
QY	64 DKWHRHRRVERIPDMAYRFLTEICGYLLPEPDDFTITTSQVDAITTTAGPQLVVPVINA 123				
DB	70 DAWYKRNKQSGYQADYQQLXIDIGLPEGAGSVSTINVDPEITHIAGPQLVVPVINA 129				
QY	124 RFALNAANRWGSLDYALYGTDTVIBETGAKRGPTYNKRGDKVIARVAFKLDSDVPLSS 183				
DB	130 RYALNAANRWGSLDYALYGTDTVISEADGAEKGGKPNKRGKVIARVAFKLDSDVPLSS 189				
QY	184 GSGFGNATGTGTVDGQGVNVALPKS-TGLANPCGPAGYTGAEPSPTVLLNHGHLIEILI 242				

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).

DR EMBL; AE014458; AAN30550.1; --
 DR TIGR; BR1648; --
 DR HAMAP; MF 00641; --
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR006253; Malate synthase.
 DR Pfam; PF01274; Malate synthase; 1.
 DR TIGRFAMs; TIGR01345; malate syn G; 1.
 DR Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
 KW Complete proteome.
 FT ACT_SITE 345 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 636 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 728 AA; 79966 MW; F95869D002A14EDE CRC64;

Query Match 60.7%; Score 2313; DB 1; Length 728;
 Best Local Similarity 61.6%; Pred. No. 1.5e-143;
 Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 4 VSVGNLRIARVLYDFVNNEALPGTDIDPDSFWAGVKVADLTQNALNARDELQAI 63
 Db 10 VEIEGLAVAPELVEFLAKEAPGTGVEPEKFKGFAIIRDLAPKRALIAKDELQARI 69
 QY 64 DKWHRVRVIEPIDMDAYRQFLTEIGYLLPEDDFTITTSVDAEIITTAGPQLVVPVNA 123
 Db 70 DANYKENRDKGYQADYQQLKIDIGYLLPEGGAFSTTNVDPEITHIAGPQLVVPVNA 129
 QY 124 RFALNAANRWGLSLDALYGTVDVTPETDGAEGKPTYNKVRGDXVIAYARKFLDSDVPLSS 183
 Db 130 RYALNAANRWGLSLDALYGTDAISEADGAEGKGNPKRGEKVIAWAKNFLDESAPLST 189
 QY 184 GSGDGTGFTVQDQGVVALPDKS-TGLANFGFAGYTGNAESPTSVLLINHLHIEILI 242
 Db 190 GKWADVAGLVNDGKLEIRITDGSATLKDDESFGKYGDAASPTVLLAKNMHVDIVI 249
 QY 243 DPESQVGTDRAGVKDVIKLESAITTIMDPEDSVAADAAKVLGYRNWLNKGLDAAAV 302
 Db 250 NADHPKGTDPADHADVLESALSTIODCEDSAVDAEDKVAVYRNWLNKGLDFTF 309
 QY 303 KDKGTAFRLVNRDRNYTAPGGQFTLPGRSLNFVRNVLMTNDAIVTDGSEVPEGIM 362
 Db 310 EKNGKQWTRNLGDRYTPDGSSTLTKGHSLMLVRNVLMTNPAILDAGNEVPEGIM 369
 QY 363 DALFTGLIATHGKASDVNGPLNSRTGSIYIKPKMHGPAEVAFTCELSFVEDVIGUP 422
 Db 370 DAFTSLIALHDI---GPNGRHNSREGSVYIVKPMHGFEEVAFANEFTITEMLGK 426
 QY 423 QNTMKIGIMDEERTTNWKACIAAADRVVFTINTGFLDRTGDEIHTSMEAGPMVRKGTM 482
 Db 427 PNTLKIGIMDEERTTNLKEAIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 486
 QY 483 KSQPWILAYEDHNVDAGLAAGFSGRAOVGKGMWMTLXADNVETKIAOPRAGASTAWVP 542
 Db 487 KQANWIGAVEQWVDIGLEGLSGHAQIGKGMWMPDMWMAAMLEQKIAHPRAGASTAWVP 546
 QY 543 SPTAATLHLYHQVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDNCCQ 602
 Db 547 SPTAATLHLYHQVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDNCCQ 606
 QY 603 SILGYVVRVWQGVGSKVPDIHDVALMEDRATLRISQALLANLWHLGHVITSADVRASLE 662
 Db 607 GILGYVVRVWQGVGSKVPDIHDVALMEDRATLRISQALLANLWHLGHVITSADVRASLE 666
 QY 663 RMAPLVDRQVAGVYRVPAPNPDSDIAFLAQELISGAQOPNGVTEPILHRRREFK 721
 Db 667 RMAAIVDKQNEGDPLRPMADFDKSIATQACCDLVFKGEEQNGVTEPILHRRREFK 725

RESULT 14
 MASZ_CORGL
 ID MASZ_CORGL STANDARD; PRT; 738 AA.

P42450;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR ACEB OR GCL329.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI TaxID=1718;
 RN [1] SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RC MEDLINE=95111631; PubMed=7812449;
 RA Reinscheid D.J., Eikmanns B.J., Sahm H.;
 RT "Malate synthase from Corynebacterium glutamicum: sequence analysis
 of the gene and biochemical characterization of the enzyme.";
 RL Microbiology 140:3099-3108(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=ATCC 13059 / AS019;
 RC Lee H.S., Sinskey A.J.;
 RT "Molecular characterization of aceB, a gene encoding malate synthase
 in Corynebacterium glutamicum";
 RL J. Microbiol. Biotechnol. 4:256-263(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RC Nakagawa S.;
 RA "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
 CC CoA.
 CC -!- ENZYME REGULATION: Inhibited by oxalate, glycolate and ATP.
 CC -!- PATHWAY: Glyoxylate bypass; second step.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the malate synthase family. GICB subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X78491; CAA55243.1; --
 DR EMBL; L27123; AAA68074.1; --
 DR EMBL; AP005281; BAB99722.1; --
 DR PIR; I40715; I40715.
 DR HSP; P37330; LD8C.
 DR HAMAP; MF 00641; --
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR006253; Malate synthase.
 DR Pfam; PF01274; Malate synthase; 1.
 DR TIGRFAMs; TIGR01345; malate syn G; 1.
 DR Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
 KW Complete proteome.
 FT INIT_MET 0
 FT ACT_SITE 355 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 646 CATALYTIC ACID (BY SIMILARITY).
 SQ SEQUENCE 738 AA; 82231 MW; 72AA0663AE7C87F4 CRC64;

Query Match 60.1%; Score 2289; DB 1; Length 738;
 Best Local Similarity 61.3%; Pred. No. 5.8e-142;
 Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

QY 1 TDRVSGNLRIRVLYDFVNNEALPGTDIDPDSFWAGVKVADLTQNALNARDELQ 60
 Db 19 TERVDAGGMQVAKVLYDFVTEAVLPRVGVDAEKWSGFAALARDLTFRNRELLARRDELQ 78
 QY 61 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEDDFTITTSVDAEIITTAGPQLVVPV 120

DB 79 MLIDYHNN-SGIDQEAIVEDFLKEIGYLVEPEAAEINTQNVDTBISTAGPOLVPI 137
QY 121 LNARFALNAANRWGSYDLYGTDVTPETDGAEGKTYNNKRGDKVIAVARFLDSDVP 180
DB 138 LNARFALNAANRWGSYDLYGTDVTPETDGAEGKTYNNKRGDKVIAVARFLDSDVP 197
QY 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPCQFAGYTCGAESPVSLLINHLHIEI 240
DB 198 LDGASHADVKEYNITDGLAAHIGDSVRLKNRESYRGFTGNFLDPAIFLLENGHLIEL 257
QY 241 LIDPESQVGTTRDAGVKDVLIESAITTIMPDSVAADAAKVLGYRNWGLNKGDLAA 300
DB 258 QIDPWHPIGKADKTGLKDVLESAITTIMPDSVAADAAEDKTGLYSNNFGLNTGELKE 317
QY 301 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGSLMFRVNVGHLMTNDALIVDTGSEVFE 360
DB 318 EMSKNGRIFTBELAKDRYIIGRNGTELVHGRSLFLVRNVGHLNQNPSIL-IDGEEIFEG 376
QY 361 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGPFAVFTCELFGRVEDVLG 420
DB 377 IMDAVLTVCAPIGLAPON---KXNSRKGSIYIVKPKMGHGPFAVFTCELFGRVEDVLG 433
QY 421 LPQNTMKIGIMDEBERRTTVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 480
DB 434 LPRHTLKVGVMDEBERRTSVNLDAIMEVADRLAFINTGFLDRTGDEIHTSMEAGPMVRKA 493
QY 481 TMKSQPMILAYEDHNVDAAGLAAGFSGRAQVCGKGMWMTLMADVETKIAQPRAGASTAW 540
DB 494 DMQTAPEWQAVENNVNDAGIORGLPGKQIIGKGMWMTLMADVETKIAQPRAGASTAW 553
QY 541 VPSPTAATLHALHYQVDVAQVQGL-AGKRRTIEQLTIPLAKELAWAPDEIREVDN 599
DB 554 VPSPTGATLHATHVHLVDVFKVQDELRAAGRRDRLNLTPTAPNTNWSSEKKEMDN 613
QY 600 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 659
DB 614 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 673
QY 660 SLERMAPLVDRQNAQGVAYRMAPNFDDSIAPLAAQELILSGAQQPNGYTEPILHRRRE 719
DB 674 SLERMAPLVDRQNAQGVAYRMAPNFDDSIAPLAAQELILSGAQQPNGYTEPILHRRRE 733
QY 720 FKAR 723
DB 734 FKAR 737

RESULT 15

ID MASZ_COREF STANDARD; PRT; 748 AA.
AC Q8FNB3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASZ OR CE2231.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. G1CB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005221; BAC19041.1; -.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn_G; 1.
KW Transferase; glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 362 362 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 653 653 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;

Query Match 59.7%; Score 2273; DB 1; Length 748;

Best Local Similarity 61.5%; Pred No. 6.5e-141;

Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;

QY 1 TDRSVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 60

DB 26 TERVTGGMQVAKVLRDLFTESVLPBGVDAERFWNGFGDIVRDMTPRRELLARDELQ 85

QY 61 AQIDKWHRRVTEPDMDAYRQFLTEIGVLLPEPDDFTTTSGVDABITTTAGPOLVVPV 120

DB 86 AQIDSEYRENPKP-DPEKYEAFLEIGVLYDEPAPAIRTONIDSEIATTAGPOLVVPV 144

QY 121 LNARFALNAANRWGSYDLYGTDVTPETDGAEGKTYNNKRGDKVIAVARFLDSDVP 180

DB 145 LNARFALNAANRWGSYDLYGTDVTPETDGAEGKTYNNKRGDKVIAVARFLDSDVP 204

QY 181 LSSGSGDATGFTVQDGLVVALPKSTGLANPCQFAGYTCGAESPVSLLINHLHIEI 240

DB 205 LDGASHADVKEYNITDGLAAHIGDSVRLKNRESYRGFTGNFLDPAIFLLENGHLIEL 264

QY 241 LIDPESQVGTTRDAGVKDVLIESAITTIMPDSVAADAAKVLGYRNWGLNKGDLAA 300

DB 265 QIDPWHPIGKADKTGLKDVLESAITTIMPDSVAADAAEDKTGLYSNNFGLNTGELKE 324

QY 301 AVDKDGTAFRLVLRNDRNYTAPGGQFTLPGSLMFRVNVGHLMTNDALIVDTGSEVFE 360

DB 325 EVAKGRDTRFTRKLNDRVDFVIGKNGAELTLHGSRLLFVRNVGHLMTNPAIL-VDGSEIYEG 383

QY 361 IMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGPFAVFTCELFGRVEDVLG 420

DB 384 IMDAITTVCAPIGLAPONKK---NSRKGSIYIVKPKMGHGPFAVFTCELFGRVEDVLG 440

QY 421 LPQNTMKIGIMDEBERRTTVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKG 480

DB 441 LPRHTLKVGVMDEBERRTSVNLDAIMEVADRLAFINTGFLDRTGDEIHTSMEAGPMVRKA 500

QY 481 TMKSQPMILAYEDHNVDAAGLAAGFSGRAQVCGKGMWMTLMADVETKIAQPRAGASTAW 540

DB 501 DMQTAPEWQAVENNVNDAGIORGLPGKQIIGKGMWMTLMADVETKIAQPRAGASTAW 560

QY 541 VPSPTAATLHALHYQVDVAQVQGL-AGKRRTIEQLTIPLAKELAWAPDEIREVDN 599

DB 561 VPSPTGATLHATHVHLVDVFKVQDELRAAGRRDRLNLTPTAPNTNWSSEKKEELDN 620

QY 600 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 659

DB 621 NCOSILGYVVRWVDOGVGCSKVPDTHDVALMEDRATLRISQLLANLHGHVITSADVRA 680

QY 660 SLERMAPLVDRQNAQGVAYRMAPNFDDSIAPLAAQELILSGAQQPNGYTEPILHRRRE 719

Db 681 SLERMAVVVEQVAGDPNYLWAPNFTESVAFQAAEDLILKGTSPAGYTEPILHARRRE 740
QY 720 FK 721
Db 741 FK 742

Search completed: November 21, 2003, 16:04:15
Job time : 11.7256 secs

This Page Blank (uspio)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 36.4127 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-2
Perfect score: 3810
Sequence: 1 TDRVSGNLRIRARLVDFVN.....KARAEKAPSDRAGDDAAR 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2634	69.1	724	2 Q9AE55	Q9AE55 rhodococcus
2	2569	67.4	725	16 Q91636	Q91636 pseudomonas
3	2416	63.4	721	16 Q98DK4	Q98DK4 rhizobium 1
4	2374.5	62.3	744	16 Q8UJ85	Q8UJ85 agrobacteri
5	2367.5	62.1	723	2 Q937W7	Q937W7 rhizobium 1
6	2351.5	61.7	723	16 Q92TA4	Q92TA4 rhizobium m
7	2318	60.8	728	16 Q8YIR3	Q8YIR3 brucella me
8	2313	60.7	728	16 Q8FZ50	Q8FZ50 brucella su
9	2273	59.7	748	16 Q8FNB3	Q8FNB3 corynebacte
10	2256.5	59.2	727	16 Q8XB03	Q8XB03 bacillus ha
11	2168.5	56.9	723	16 Q8FDN6	Q8FDN6 escherichia
12	279.5	7.3	538	16 Q8FQR6	Q8FQR6 xanthomonas
13	258	6.8	540	16 Q8PDU8	Q8PDU8 xanthomonas
14	206	5.4	907	5 Q9U3Q5	Q9U3Q5 caenorhabdi
15	198.5	5.2	968	5 O17353	O17353 caenorhabdi
16	190.5	5.0	533	16 Q8X609	Q8X609 escherichia

ALIGNMENTS

RESULT 1

Q9AE55 PRELIMINARY; PRT; 724 AA.
ID Q9AE55;
AC Q9AE55;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Malate synthase.
GN VICA.
OS Rhodococcus fascians.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1828;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DI88;
RA Vereecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
RA Goethals K.;
RT "Characterization of a chromosomal locus that affects pathogenicity in
RT Rhodococcus fascians";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301559; CAC35701.1; -;
DR HSSP; P37330; ID8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthG.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn.G; 1.
SQ SEQUENCE 724 AA; 78609 MW; F889F883890995E CRC64;

Query Match 69.1%; Score 2634; DB 2; Length 724;
Best Local Similarity 69.0%; Pred. No. 2.2e-162;
Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;

QY 1 TDRVSGNLRIRARLVDFVNALPGTDIDPSFWAGVKVADLTQNALNARDELQ 60
DB 2 TDRVQAGLQVAKVLFDFVEKEALPGTDLSEAFWAGASVIADLAFNKKALLAVRDIIQ 61
QY 61 AQIDKWHRRRVIEPIEDMDAYRQFLTEIGYLLPEPDDFTITTSVGDAEITTTAGPQLVVPV 120

17 190 5.0 533 16 Q8ZIW0
18 187 4.9 536 16 Q8ZKL4
19 187 4.9 556 16 Q9KTZ6
20 186.5 4.9 539 16 Q8FB63
21 184 4.8 543 5 Q8MYJ5
22 182.5 4.8 551 3 Q9P8Q2
23 181.5 4.8 551 16 Q8EF8
24 179.5 4.7 530 16 Q93J61
25 179 4.7 545 16 Q8DEX5
26 177 4.6 188 16 Q8D3P1
27 175.5 4.6 525 16 Q9A7F8
28 160 4.2 562 10 Q9LZC3
29 154.5 4.1 202 16 Q9KKZ3
30 154.5 4.1 529 16 Q8YZN2
31 154 4.0 540 16 Q9RKU9
32 149 3.9 532 16 Q8ZAR5
33 149 3.9 543 16 Q8DIU4
34 149 3.9 549 16 Q8EGV8
35 147 3.9 540 2 Q9ALX7
36 143.5 3.8 542 3 Q93857
37 142 3.7 528 2 Q8VM95
38 142 3.7 556 10 Q9AXL7
39 140 3.7 813 17 Q8ZXH2
40 139 3.6 3455 16 Q9P9U6
41 138.5 3.6 531 16 Q8ENS2
42 137.5 3.6 4848 2 O07944
43 135.5 3.6 521 16 Q9RYN3
44 135.5 3.6 3442 16 Q9PBE8
45 135 3.5 3816 2 Q9KIV3

Q8ZIW0 salmonella
Q8ZKL4 salmonella
Q9KTZ6 vibrio chol
Q8FB63 escherichia
Q8MYJ5 dictyosteli
Q9P8Q2 candida alb
Q8EF8 shewanella
Q93J61 streptomyce
Q8DEX5 vibrio vuln
Q8D3P1 vibrio vuln
Q9A7F8 caulobacter
Q9LZC3 arabidopsis
Q9KKZ3 vibrio chol
Q8YZN2 ralstonia s
Q9RKU9 streptomyce
Q8ZAR5 yersinia pe
Q8DIU4 yersinia pe
Q8EGV8 shewanella
Q9ALX7 streptomyce
Q93857 laccaria bi
Q8VM95 alcaligenes
Q9AXL7 musa acumin
Q8ZXH2 xylobaculum
Q9P9U6 xylella fas
Q8ENS2 oceanobacil
O07944 streptomyce
Q9RYN3 deinococcus
Q9PBE8 xylella fas
Q9KIV3 streptomyce

Db 62 GKVDAAHGEHAGAEYDRAAYKAFLEIGYLLDEPADFQIHTSGVDTETITTAGPQLVVPV 121
QY 121 LNARFALNANARWGLSYDALYGTDTVPETDGAESKGYTNKVRGDKVIYARFELDSVP 180
Db 122 LNARFALNANARWGLSYDALYGTDAIPETDGAESKGYTNKVRGDKVIYARFELDSVP 181
QY 181 LSSGSGFATGFTVQDQGLVVALPKST-GLANPGQFAGYTGAAESPTSVLLINHGHLIE 239
Db 182 LSSGSHVCTTGVVDAAASLTVLADGSIYVGLKDSQLLQVGTDPADTAIFVHNGHLFE 241
QY 240 ILIDPESQVGTDRAGVKDVILESAITIMPFDSVAADAAVKVLYRNWGLKNGDLA 299
Db 242 IQIDPESPIGKTDGAGVKDVILESAITIMPFDSVAADAAVKVLYRNWGLKNGDLT 301
QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHMTNDAIVDTDGSVEFE 359
Db 302 EVSKGGKTFTRANKKORTYTSVDGSELTLLHGRSLLFVRNVGHMTNDAILDADGNEVE 361
QY 360 GIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSFSEV 419
Db 362 GILDALFTSLAGLSLTDPNV--LSNSRTGSLYIVKPKHGPDEVAFTAELFGRVEQVL 418
QY 420 GLPQNTWKIGIMDEERTTNVNLKACIKAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 479
Db 419 GLPNTLKVGMDEERTTNVNLKACIKAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 478
QY 480 GTMKSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTMTLMADWVETKIAQPRAGASTA 539
Db 479 GAMKGEKWIAYEDFNVDVTLGAGLQKQKAIKKGWAWPDLMDMLQKICHKPKAGANTA 538
QY 540 WVPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELANWAPDEIREVDN 599
Db 539 WVPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELANWAPDEIREVDN 598
QY 600 NCQSILGVVWVWQVGVCSKVPDIHDVLMEDRATIRISSQLLANWLRHGVITTSADV 659
Db 599 NSQSILGVVWVWQVGVCSKVPDIHDVLMEDRATIRISSQLLANWLRHGVITTSADV 658
QY 660 SLERMAPLVDRQAGDVAYRMAPNFDDSIAPLAAQELILSGAQPNGYTEPIHLRRRE 719
Db 659 SLERMAPLVDRQAGDVAYRMAPNFDDSIAPLAAQELILSGAQPNGYTEPIHLRRRE 718
QY 720 FKA 722
Db 719 YKA 721

RESULT 2

Q91636
ID Q91636 PRELIMINARY; PRT; 725 AA.
AC Q91636
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.B., Hunsigle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
[1]
RN NCBI_TaxID=287;
OC Pseudomonadaceae; Pseudomonas.
OS *Pseudomonas aeruginosa*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
RN NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.B., Hunsigle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
[1]

DR EMBL; AB004485; AAG03871.1; -
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
KW Complete proteome.
SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;
Query Match 67.4%; Score 2569; DB 16; Length 725;
Best Local Similarity 69.1%; Pred. No. 3.7e-158;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
QY 1 TDRVSVGLNRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTTPONQALLNARDLQ 60
Db 2 TERVQVGLQVAKVLFDFVNEALPGTGVASDTFTWGAENVINDLAPKNKALLAKRDLQ 61
QY 61 AQIDKWHRRRVIEPTDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVVPV 120
Db 62 AXIDGWHQARQAQAHDAVAYKAFLEIEIGYLLPEAEDFOAGTQNVDDDEIARMAGPQLVVPV 121
QY 121 LNARFALNANARWGLSYDALYGTDTVPETDGAESKGYTNKVRGDKVIYARFELDSVP 180
Db 122 LNARFALNANARWGLSYDALYGTDTVPETDGAESKGYTNKVRGDKVIYARFELDSVP 181
QY 181 LSSGSGFATGFTVQDQGLVVALPKST-GLANPGQFAGYTGAAESPTSVLLINHGHLIE 239
Db 182 LSSGSHVCTTGVVDAAASLTVLADGSIYVGLKDSQLLQVGTDPADTAIFVHNGHLFE 241
QY 240 ILIDPESQVGTDRAGVKDVILESAITIMPFDSVAADAAVKVLYRNWGLKNGDLA 299
Db 242 IQIDPESPIGKTDGAGVKDVILESAITIMPFDSVAADAAVKVLYRNWGLKNGDLA 301
QY 300 AAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHMTNDAIVDTDGSVEFE 359
Db 302 EVSKGGKTFTRANKKORTYTSVDGSELTLLHGRSLLFVRNVGHMTNDAILDADGNEVE 361
QY 360 GIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAEVAFTCELSFSEV 419
Db 362 GILDALFTSLAGLSLTDPNV--LSNSRTGSLYIVKPKHGPDEVAFTAELFGRVEQVL 418
QY 420 GLPQNTWKIGIMDEERTTNVNLKACIKAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 479
Db 419 GLPNTLKVGMDEERTTNVNLKACIKAAADRVVFINTEGLDRTGDEBIHTSMEAGPV 478
QY 480 GTMKSQPMILAYEDHNVDAAGLAFSGRAQVKGKMTMTLMADWVETKIAQPRAGASTA 539
Db 479 GAMKGEKWIAYEDFNVDVTLGAGLQKQKAIKKGWAWPDLMDMLQKICHKPKAGANTA 538
QY 540 WVPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELANWAPDEIREVDN 599
Db 539 WVPSPATAATHALHYHVDVAAVQOGLAGKRATIEQLLTIPLAKELANWAPDEIREVDN 598
QY 600 NCQSILGVVWVWQVGVCSKVPDIHDVLMEDRATIRISSQLLANWLRHGVITTSADV 659
Db 599 NSQSILGVVWVWQVGVCSKVPDIHDVLMEDRATIRISSQLLANWLRHGVITTSADV 658
QY 660 SLERMAPLVDRQAGDVAYRMAPNFDDSIAPLAAQELILSGAQPNGYTEPIHLRRRE 719
Db 659 SLERMAPLVDRQAGDVAYRMAPNFDDSIAPLAAQELILSGAQPNGYTEPIHLRRRE 718
QY 720 FKA 722
Db 719 YKA 721
RESULT 2
Q91636
ID Q91636 PRELIMINARY; PRT; 725 AA.
AC Q91636
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.B., Hunsigle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
[1]
RN NCBI_TaxID=287;
OC Pseudomonadaceae; Pseudomonas.
OS *Pseudomonas aeruginosa*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
RN NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.B., Hunsigle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
[1]

DE Malate synthase G.
GN MR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
ON NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003004; BAB51267.1; -
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_synG; 1.
KW Complete proteome.
SQ SEQUENCE 721 AA; 78058 MW; 65376311A7E1BFDF CRC64;
Query Match 63.4%; Score 2416; DB 16; Length 721;
Best Local Similarity 64.4%; Pred. No. 3.1e-148;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;
QY 1 TDRVSGNLRIRARLVDFYNNALPGTDIDPSPWAGVQKVVADLTQNALNARDLQ 60
DB 2 TDRIETAGLRIGELHDFVAGEALPGTGAADAFWSGSAIVHDLAPKRALKKRDAMQ 61
QY 61 AQIDKWHRRVPEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTAGPOLVVPV 120
DB 62 ERLDGWYRNG-APVDMVEYKFLKEIGYLVPEGPAFVSVDNDVPEIAWAGPOLVPEV 120
QY 121 LNARFALNAANRWGSLYDALYGTVDVPEIDGAEKGYTYKVRGDKVIAYARKFLDDSV 180
DB 121 MARYALNAANRWGSLYDALYGTDAIPETGAEGKGNPARGAKVIANAKDFLDQVP 180
QY 181 LSGSGFDATGTFVQDQGLVWALPDKSTGLANPGPAGVTGAESFTSVLLNHGLHIEI 240
DB 181 LTSKGWAGVNGLSVAHGALKGLGAGAGTTLADPRFAGYRGAANPDVALLVNGHGLHIEI 240
QY 241 LIDPESQVGTDRAGVKVILSAITTTIDFEDSVAADAAKVLGYRWLGNLKGDLAA 300
DB 241 VDRNNQIGRTDAGTADVILSALTITIQCDSDVAANDAQKVVYRWNLGMLKGDIAE 300
QY 301 AVDKGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNTDAIVDTGSEVFEG 360
DB 301 EISKAGRSFVRKLNADRAVYAPAGGQITVPGRSMLVRNVGHLMNTNPAILDRDGNVEPEG 360
QY 361 IMDALFTGLIAHGLKASDV--NGPLNSRTSGIYVKKPMGHPAEVATCFLFRVEDV 418
DB 361 IMDAALTALIALH-----DVGPEGRNRSRAGSMYVVKPMGHPAEVAFVIFRVEAL 415
QY 419 LGLPONTMKGIMDEERRITVNLKACIKAAADRVFINTGFLDRDTDEHTSMEAGPMVR 478
DB 416 LGMPNTIKGIMDEERRITVNLKEAIRAREVVFINTGFLDRDTDEHTSMEAGPMIR 475
QY 479 KGMKKSQPMILAVEDHNVNAGLAAGFSGRAQVKGKMTWTELMADNVETKIQAPRAGAST 538
DB 476 KGMKKAAMTISAYEAWNVDTGLECGLAGHAQIKGNWAMPDLMAAMLEQKIAPHRAGANT 535
QY 539 AWVPSPTAATLHLYHVDVVAQVQGLAGKRRATIEQLTTPLAKELAWAPDEIREVD 598
DB 536 AWVPSPTAATLHLYHVDVVAQVQGLAGKRRATIEQLTTPLAKELAWAPDEIREVD 595
QY 599 NNCQSILGYVVRWVDQGVCSKVPDTHDVALMEDRATLRISQLLANWLHGHVITSADVR 658
DB 596 NNAQGLILGYVVRWVDQGVCSKVPDTHDVALMEDRATLRISQLLANWLHGHVITSADVR 655

QY 659 ASLERMAPLVDRONAGVAVRVPAPNPDDSIATFAAQELILSGAQPNQNGVTEPILHRRR 718
DB 656 DSLQMAAIVDRQNVGDPYRVPAPDFDKSIATFAQACDLVFKGTQSNQNGVTEPVLHARR 715
QY 719 EFKAR 723
DB 716 ELKQA 720
RESULT 4
Q8UJ85 PRELIMINARY; PRT; 744 AA.
AC Q8UJ85
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase G.
GN GLCB OR ATU0047 OR AGR_C 78.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr., Grant C.,
RA Chapman P., Devening J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tinsley S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestor E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Houriello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houriello B., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE008979; AAL41078.1; -
DR EMBL; AE007947; AAK85871.1; -
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_synG; 1.
KW Complete proteome.
SQ SEQUENCE 744 AA; 80991 MW; FB74ED358D650319 CRC64;
Query Match 62.3%; Score 2374.5; DB 16; Length 744;
Best Local Similarity 62.3%; Pred. No. 1.6e-145;
Matches 462; Conservative 97; Mismatches 162; Indels 21; Gaps 4;
QY 1 TDRVSGNLRIRARLVDFYNNALPGTDIDPSPWAGVQKVVADLTQNALNARDLQ 43
DB 6 SDRLSVQSMPEYKEAHVSRTRDKFGLSIDDRLYAFLTDEVLPGTGLDSETFEGFSAIVH 65
QY 44 DLTQNALNARDLQADIKWHRRVPEIDMDAYRQFLTEIGYLLPEPDDFTITTS 103
DB 66 ELSPKRELLAKRALQEKIDGWYRENG-APSDFAVEAFLEIGYLLPEPGFKVETNN 124
QY 104 VDAETTTTAGTQVLPVNLNARFALNAANRWGSLYDALYGTVDVPEIDGAEKGYTYKVR 163

```

125 VDBIAVAGPQLVVPVNNARYALNAANARWSLYDALYGTDAISDAGAEKGRGNPKR 184
164 GDKYIARKEKLDSDVPLSSGSGFDATGFTVDDQQLVVALPDKSTGLANPQAGYTGAA 223
185 GDKYIARKEKLDSDVPLSSGSGFDATGFTVDDQQLVVALPDKSTGLANPQAGYTGAA 244
224 ESPTSVLLINHLGLHIEILIDPESQVGTTRAGVNDVILESAITTIMDFEDSVAADAAK 283
245 AKPATILGKNGLHTEIVDPSTIGKSDRAGISDVILESAITTIMDFEDSVAADAAK 304
284 VLGRNMLGKNGGLAAAVDKGTAFILRNDRNYTAPGGQFTPLPGRSLMFRVNVGHL 343
305 VLVTGNMLGMRGDLTEAVSGKNTFTLRNLPDRYYPADGSAITLPLGRSLMFRVNVGHL 364
344 MTNDALVDDTSGSEVFEIGMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGPA 403
365 MTNPAILDRDQDVEGIMDAVVTALIALYDVGPS---GRQNSRAGSMVIVKPKMHGPE 421
404 EVATCELFSEVDEVLGPQNTKIGIMDEERTTVNLKACIKAAADRVVFINTEGLDRT 463
422 EVAFANEIARVENLVGAPNTKMGIMDEERTTVNLKESIRAAKDRVVFINTGFLDRT 481
464 GDEHTSMEAGPMYRKCTKMSQPIILAYEDHNVDAGLAAGSGRAQVKGKMTMTTELAD 523
482 GDEHTSMEAGPMYRKCTKMSQPIILAYEDHNVDAGLAAGSGRAQVKGKMTMTTELAD 541
524 NVETKIAOPRAGASTAWPSPATAATLHALHYHOVDVAQVQGLAGKRATIEQLLTIPLA 583
542 MLEQKIAHPKAGANTAWPSPATAATLHALHYHOVDVAQVQGLAGKRATIEQLLTIPLA 601
584 KELAWEADBEIREVDNQCOSILGVVVRWDQVGCCKVPDIHVALMEDRATIRISSQLL 643
602 PRPNTPEEIQRELDNNAQGLGVVVRWDQVGCCKVPDIHVALMEDRATIRISSQLL 661
644 ANWLHRGVITSADVRASLERMAPLVDQNGADVAIRPMAFNFDSDIAFLAAQELISGAQ 703
662 ANWLHRGVITSADVRASLERMAPLVDQNGADVAIRPMAFNFDSDIAFLAAQELISGAQ 721
704 QPNGYTEPILHRRRRREFKARA 725
722 QPNGYTEPILHRRRRREFKARA 743

RESULT 5
Q937W7 PRELIMINARY; PRT; 723 AA.
ID Q937W7;
AC Q937W7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase G.
GN MSG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V939;
RA Garcia de los Santos A., Hynes M.F.;
RL "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059637; AAL17965.1; -
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;

```

Query March 62.1%; Score 2367.5; DB 2; Length 723;
 Best Local Similarity 64.1%; Pred. No. 4.4e-145;
 Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;

```

QY 3 RYSGNLRIRVLYDFVNEALPGTDIDPDSFMAGVDKVVADLTPQNALNARDELQAO 62
DB 3 RYSGNLRIRVLYDFVNEALPGTDIDPDSFMAGVDKVVADLTPQNALNARDELQAO 62
QY 63 IDKWHRRVIFIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPVNL 122
DB 63 IDKWHRRVIFIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPVNL 122
QY 123 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 181
DB 123 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 181
QY 181 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 241
DB 181 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 241
QY 241 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 301
DB 241 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 301
QY 301 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 361
DB 301 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 361
QY 361 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 421
DB 361 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 421
QY 421 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 481
DB 421 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 481
QY 481 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 541
DB 481 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 541
QY 541 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 601
DB 541 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 601
QY 601 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 661
DB 601 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 661
QY 661 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 721
DB 661 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 721
QY 721 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 771
DB 721 ARFALNANARWSLYDALYGTDAIPESDGAEGKSNPKRGEKVIARVRFLOTSAPLQ 771

```

RESULT 6

Q92TA4

ID Q92TA4 PRELIMINARY; PRT; 723 AA.

AC Q92TA4;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Probable malate synthase G protein (BC 4.1.3.2).

GN GLCB OR R00062 OR SMC02581.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,

RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

Db 547 SPTAATLHATHYKHIDVAAQVKLSRPRAKLDDILSPVAVRPNTDPDDIQHEIDNNAQ 606
 QY 603 SILGVVVRWDQGVGCSKVPDIHVALMEDRATLRISQLLANWLRHGVITSADVRASLE 662
 Db 607 GILGVVVRWDQGVGCSKVPDINNVLGMDRATLRISAQHIANWLVHGVVSEAQVMTMK 666
 QY 663 RMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQBELILSGAQPNGYTEPIHLRRRREFK 721
 Db 667 RMAAIVDKQNEGDPDLYRPMADPKSIATFOACDLVFKGREQPNGYTEPVLHRRRLLEK 725

RESULT 8
 Q8FZ50
 ID Q8FZ50 PRELIMINARY; PRT; 728 AA.
 AC Q8FZ50;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Malate synthase G.
 GN GLCB OR BR1648.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Unayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraut R., Shetty J., Malek J., van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.B., Lindler L.E., Helling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
 DR EMBL; AE014458; AAN30550.1; --
 DR TIGR; BR1648; --
 KW Complete proteome.
 SQ SEQUENCE 728 AA; 79966 MW; F35669D002A14EDE CRC64;

Query Match 60.7%; Score 2313; DB 16; Length 728;
 Best Local Similarity 61.6%; Pred. No. 1, 5e-141;
 Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 4 VSGNLRIRARVLYDFVNEALPGTIDPDSFWAGVDKVVADLTQNOALLNARDELQAI 63
 Db 10 VEIEGLAVAPLVEFLAKAEPGTGVEPEKFWGFAAIIRDLPKRRALLAKDELQARI 69
 QY 64 DKWRRRRIEIPIDMDAYRQFLTEIGYLLPSPDDPTITTSVGDAEITTTAGPQLVVPVNA 123
 Db 70 DAWTKERDKGYSQADYQOFLKIDIGYLLPEGGAFSVSTINVDPEITHIAGPQLVVPVNA 129
 QY 124 RFALNAANARWGLSYDALYGTDPVPEIDGAEKGYTNKVGDKVIAYARKFLDSDVPLSS 183
 Db 130 RYALNAANARWGLSYDALYGTDAISEADGAEKGYNPKGEKVIYAWAKNFDLDESAPLT 189
 QY 184 GSPGDATGFTVODGQLVVALPKDS-TGLANPGQFAGYTGAEPTSVLINHLHGLIEILI 242
 Db 190 GKWADVAGLVNDGKLEIRTDGATLTDKDSQFNGDAAPTNVLAKHNMHVDIVI 249
 QY 243 DPESQVETTRAGVKVDVILSAITITIMDFDSVAADDAADKVLGYRNWLNGLNGDLAAAV 302
 Db 250 NADHPICKTDPAHADVVLESAISTIODCEDSIAADVAEDKVAAYRNWLNGLNGKLEDITF 309
 QY 303 DKDGTAFELVRNDRNTAPGGGFTLPGRSLFVRNVGHMTNDALVDTGSEVEPEGIM 362
 Db 310 EKNQKQMTRRNGDRNTTAPDGSFTLTGKHSMLVRNVGHMTNPAILDAGNEVEPEGIM 369
 QY 363 DALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLGLP 422
 Db 370 DAAFTSLIALHDI--GPNGRHNSRSGSIYIVKPKMHGPEEVAFANEIFTRTEELGMK 426

QY 423 QNTMKIGIMDEBERRTTNVLKACIKAAADRVVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 482
 Db 427 PNTLKIGIMDEBERRTTNVLKAIKAAADRVVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 486
 QY 483 KSQPMILAYEDHNVDAGLAAGFSQRAQVKGKMTWMTLMADMVETKIAQPRAGASTAWVP 542
 Db 487 KQAAAMIGAVEQNWVDIGLECGLSGHAQIKGKMWAMPDMMAAAMLEQKIAHPKAGANTAWVP 546
 QY 543 SPTAATLHATHYKHIDVAAQVKLSRPRAKLDDILSPVAVRPNTDPDDIQHEIDNNAQ 606
 Db 547 SPTAATLHATHYKHIDVAAQVKLSRPRAKLDDILSPVAVRPNTDPDDIQHEIDNNAQ 606
 QY 603 SILGVVVRWDQGVGCSKVPDIHVALMEDRATLRISQLLANWLRHGVITSADVRASLE 662
 Db 607 GILGVVVRWDQGVGCSKVPDINNVLGMDRATLRISAQHIANWLVHGVVSEAQVMTMK 666
 QY 663 RMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQBELILSGAQPNGYTEPIHLRRRREFK 721
 Db 667 RMAAIVDKQNEGDPDLYRPMADPKSIATFOACDLVFKGREQPNGYTEPVLHRRRLLEK 725

RESULT 9
 Q8FNB3
 ID Q8FNB3 PRELIMINARY; PRT; 748 AA.
 AC Q8FNB3;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Malate synthase (EC 4.1.3.2).
 GN MASZ OR CE2231.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS-314 / AT 12310 / DSM 44549 / JCM 11189;
 RA Kwarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.;
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005221; BAC19041.1; --
 KW Lyase; Complete proteome.
 SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;

Query Match 59.7%; Score 2273; DB 16; Length 748;
 Best Local Similarity 61.5%; Pred. No. 6, 3e-139;
 Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;

QY 1 TDRVSGNLRIRARVLYDFVNEALPGTIDPDSFWAGVDKVVADLTQNOALLNARDELQ 60
 Db 26 TERVTVGNGQVAKVLRDLFTESVLPVRGVDAERFWNGFGDIVRDMTPNRELLARRDELQ 85
 QY 61 AQIDKWHRRRRIEIPIDMDAYRQFLTEIGYLLPSPDDPTITTSVGDAEITTTAGPQLVVPV 120
 Db 86 AQLDYYRENPCKP-DPEKYEAFLREIGYLVDEPAPAEIRTONIDSEIATTAGPQLVVPV 144
 QY 121 LNARFALNAANARWGLSYDALYGTDPVPEIDGAEKGYTNKVGDKVIAYARKFLDSDVSP 180
 Db 145 LNARFALNAANARWGLSYDALYGTNAIPDEDEGAEYFNVRQKVIQWRDFLDVLVP 204
 QY 181 LSSGSFGDATGFTVODGQLVVALPKDSTGLANPGQFAGYTGAEPTSVLINHLHGLIEI 240
 Db 205 LDGASHADVEXNITDGLAAHVNDGIYLRKDRAYLGFTGYFDPFTSILLQNGHLIEL 264
 QY 241 LIDPESQVGTTRAGVKVDVILSAITITIMDFDSVAADDAADKVLGYRNWLNGLNGDLAA 300
 Db 265 QIDPHTPIGKEDKTKDKDIIILSAITITIMDFDSVAADDAEDKTLGYRNWLNGLNGDL 324
 QY 301 AVDKDGTAFELVRNDRNTAPGGGFTLPGRSLFVRNVGHMTNDALVDTGSEVEFEG 360

Db 325 EVAKGDRTRFKLNDRVFIGKNGAELTLHGRSLLFVRNVGHLMTNPAIL-VDGEEIYEG 383

QY 361 IMDALFTGLTAIHLKASDVNGPLINSRTGSIYIVKPKHGPAPVAFTELSRVEDVLG 420

Db 384 IMDAITTVCAIPGAPQKKK--NSRKGSIYIVKPKHGPAPVAFTELSRVEDVLG 440

QY 421 LPONTMKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMRAGPMVRKG 480

Db 441 LPRHTLVGVWDEERTSVNLDACIMEVADLAFINTGFLDRTGDEIHTSMRAGPMVRKA 500

QY 481 TMSQPHILAYEDHNVDAAGLAFSGRAQVKGKGMWMTMELMADVETKIAOPRAGASTAW 540

Db 501 DMOAPKQAYEDHNVDAAGLAFSGRAQVKGKGMWMTMELMADVETKIAOPRAGASTAW 560

QY 541 VPSPTAATLHALHYHVDVAAVOGL-AGKRRATIEQLITIPAKELAWAPDEIREVDN 599

Db 561 VPSPTGATLHATHYHVDVAAVOGL-AGKRRATIEQLITIPAKELAWAPDEIREVDN 620

QY 600 NCOSILGYVVRVDDQVGGCKVPDIHDVALMEDRATLRSSQILANLWLRHGVITSADVRA 659

Db 621 NCOSILGYVVRVDDQVGGCKVPDIHDVALMEDRATLRSSQILANLWLRHGVITSADVRA 680

QY 660 SLRMAPLVDRQAGVAVRPMAPNFDSDTAFLAAQELILSGAQPNGYTERILHRRRE 719

Db 681 SLRMAPLVDRQAGVAVRPMAPNFDSDTAFLAAQELILSGAQPNGYTERILHRRRE 740

QY 720 FK 721

Db 741 FK 742

RESULT 10

Q9KB03 PRELIMINARY; PRT; 727 AA.

ID Q9KB03

AC Q9KB03

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Malate synthase.

GN B21133.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331 (2000).

DR EMBL; AP001514; BAB05852.1; -

DR HSSP; P37330; 1D8C.

DR InterPro; IPR001465; Malate synthase.

DR InterPro; IPR006253; Malate synthase.

DR Pfam; PF01274; Malate synthase; 1.

DR TIGRFAMs; TIGR01345; malate_syn; 1.

KW Complete proteome.

SQ SEQUENCE 727 AA; 81208 MW; CE18B459305EAE5 CRC64;

Query Match 59.2%; Score 2256.5; DB 16; Length 727;

Best Local Similarity 60.7%; Pred.No. 7.1e-138;

Matches 439; Conservative 95; Mismatches 184; Indels 5; Gaps 3;

QY 6 VGNLRARLVDFVNEALPGTDIDPDSFWAGYKVDVADLTQNALNARDELQAQIDK 65

Db 7 VGNLQVATKLITIEQALPGTGKIDFWSGVQVQITELMPENKMLAKREEIQATIDA 66

QY 66 WHERRVIEPDMDAYQELTEIGYLLPEPDDFTITTSYVDAITTAGQQLVVPVNLARF 125

Db 67 WHORNK-GPIDFSAYHSFLEEIGYLEPIPEHVITITTENVDDDEIAAQAQQLVVPVNNARY 125

QY 126 ALNANAENWGSILYDALYCTDVIPIETDGAEGKGTYNKVRGDKVIAYARFELDDSVPLSGS 185

Db 126 AINAANAENWGSILYDALYGSNVISEEDGCEKGTYNPKRGTKVIFPAKDFLDHTFPLTSGS 185

QY 186 FQDAFTGVQOQLVVALPD-KSTGLANPGQFAGVTGAESPTSVLLINHGHLHILIDP 244

Db 186 HHEALNAYIMDKQLVWVLESKVTFLKDETOFVGQSQGDPSSVILLHLHGLHVEIQIDA 245

QY 245 ESQVOTTRAGVKOVILLESALTITMDPDSVAAVDAADKVLGYRNWGLNKGDLAAAVDK 304

Db 246 RHPIGKSDRAKVKOIVLESALTITMDPDSVAAVDAADKVLGYRNWGLNKGDLAAAVDK 305

QY 305 DGTAFRLVNLDRNYTAPGGQFTLPGRLMFRNVGHLMNTDAIVTDGSEVEPEGINDA 364

Db 306 EGKTKRKLNEDRSYAPNGETFSLPGRSLMFRNVGHLMNTPIVIRTOSGEVEPEGIDG 365

QY 365 LFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAPVAFTELSRVEDVLGPN 424

Db 366 IVTSLIAKHDLQ--NGTFNRNSKGSYIVKPKHGPAPVAFTELSRVEDVLGPN 422

QY 425 TMKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMRAGPMVRKG 484

Db 423 TLKIGIMDEERTTVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMRAGPMVRKG 482

QY 485 QPWLAYEDHNVDAAGLAFSGRAQVKGKGMWMTMELMADVETKIAOPRAGASTAWPSP 544

Db 483 SSWSLAYERSNVAAGLTCGFGQRAQVKGKGMWMTMELMADVETKIAOPRAGASTAWPSP 542

QY 545 TAATLHALHYHVDVAAVOGL-AGKRRATIEQLITIPAKELAWAPDEIREVDNCOGI 604

Db 543 TAATLHALHYHVDVAAVOGL-AGKRRATIEQLITIPAKELAWAPDEIREVDNCOGI 602

QY 605 LGYVVRVDDQVGGCKVPDIHDVALMEDRATLRSSQILANLWLRHGVITSADVRLERM 664

Db 603 LGYVVRVDDQVGGCKVPDIHDVALMEDRATLRSSQILANLWLRHGVITSADVRLERM 662

QY 665 APLVDRQAGVAVRPMAPNFDSDTAFLAAQELILSGAQPNGYTERILHRRREFKARA 724

Db 663 AKWDEQAGDPAYRPMADNLEQSVAFQALVKGTEQPSGYTERILHRRREFKARA 722

QY 725 AEK 727

Db 723 AKE 725

RESULT 11

Q9FDN6 PRELIMINARY; PRT; 723 AA.

ID Q9FDN6

AC Q9FDN6

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Malate synthase G (EC 4.1.3.2).

GN GLCB OR C3705.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6.H1 / ATCC 700928;

RX MEDLINE=2338234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosesch P.,

RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,

RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence

RT of uropathogenic Escherichia coli."

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AEO16786; AAN82151.1; -

KW Lyase; Complete proteome.

SQ SEQUENCE 723 AA; 80440 MW; AAF740E5FE038F6F CRC64;

Query Match 56.9%; Score 2168.5; DB 16; Length 723;
 Best Local Similarity 58.4%; Pred. No. 3.6e-132;
 Matches 422; Conservative 113; Mismatches 173; Indels 15; Gaps 5;

QY 4 VSVGNLRIARVLYDFVNEALPGDIDPDSFWAGVQVADLPQOALLNARDELQAQOI 63
 DB 5 ITQRLRIDANFKRFDEEVLPGVELDAAAFWHNVDEIVHDLAPENKQLLAERDRQAAL 64

QY 64 DKWRRRVEIDMDAYRQELTEGYLLPEPDPFTTSGVDVAEITTTAGPQVWPVLANA 123
 DB 65 DEWHRNPGPVKDKAAKYSFLRELGYLVPQPDHVTVTETGIDSEITTSAGPQVWPANNA 124

QY 124 RFALNAANARWGSYDALYGTVDVTPETDGAEGKPTYNKVGDKVIAYARKFLDSDVPLSS 183
 DB 125 RYALNAANARWGSYDALYGSIIIPQEGAMVSG--YDQREGQVIAVRFLDESIPLEN 182

QY 184 GSFGDGATGFTVQDQVVALPD-KSTGLANPGFAGYTGAESPTSVLLINHGHLIEILI 242
 DB 183 GSYQDVVAFKVDKQLRIQKNGKETTTLTPAQVGVYRGDTAAPTCLLKXNGHLIELQI 242

QY 243 DPESQVGTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWLGKGLDAAAV 302
 DB 243 DANGRIGKDDSAHINDVIVEAAISTILDCEDSVAAVDAEDKILLYRNLLGMOGTLOERM 302

QY 303 DKDGTAFRLVLRNDRNTAPGGGFTLPGRLMFRVNVGHMTNDIAVDTDGSEVEPEGIM 362
 DB 303 EXNGRQIVKLRNDROVTAADGSEISLHGRSLLEFIRNVGHMTIPVINDSEGNEIPEGIL 362

QY 363 DALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLP 422
 DB 363 DGVMTGAIYDLKVKQ-----NSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLP 422

QY 423 QNTWKIGIMDEERTTNLAKICAAADRVVINTGFLDRTGDEIHTSMGAGPMV---RKG 481
 DB 417 PNTLKGIMDEERTTNLAKICAAADRVVINTGFLDRTGDEIHTSMGAGPMV---RKG 476

QY 483 KSPQWILAYEDHNDAGLAAGSGRAQVKGKGMVMTMELMADVETKIAQPRAGASTAWVP 542
 DB 477 KSTPWIKAYERNVLSGLFCGLRKAQIGKGMWAMPDLMDMYKSGDQLRAGANTAWVP 536

QY 543 SPTAATLHALHYQDVAAVQOGLA-----GKRATTIEQLITPLAKELAWADEIREEV 597
 DB 537 SPTAATLHALHYQTNVQSVQANTAEFTNAEFPEFLDLDLLTIPVAENANWSEVEIQOEL 596

QY 598 DNNQCSILGYVVRVVDQGVGSKVPDIHDVVALMEDRATLRISSQLLANWLRHGVITSADV 657
 DB 597 DNNVQILGYVVRVVEQIGGSKVPDIHNVALMEDRATLRISSQHLANWLRHGVITSADV 656

QY 658 RASLERMAPLVDQVQAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTPILHR-R 716
 DB 657 QASLENNAKVVDQVQAGVAVRPMAPNPDSDIAFLAQELILSGAQOQNGYTPILHR-R 716

QY 717 RRE 719
 DB 717 LRE 719

RESULT 12
 QBPQR6
 ID QBPQR6 PRELIMINARY; PRT; 538 AA.
 AC QBPQR6
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Malate synthase.
 GN MLS OR XAC0256.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP "SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fortighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Medeiros J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;

RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AE011650; AAM35148.1;
 DR InterPro; IPR006252; Malate_synthase.
 DR InterPro; IPR001465; Malate_synthase.
 DR Pfam; PF01274; Malate_synthase; 1.
 DR TIGRFAMs; TIGR01344; malate_syn_A; 1.
 KW Complete proteome.
 SQ SEQUENCE 538 AA; 59271 MW; 80A47F8D32B7C848 CRC64;

Query Match 7.3%; Score 279.5; DB 16; Length 538;
 Best Local Similarity 24.9%; Pred. No. 9.2e-10; Indels 107; Gaps 20;
 Matches 123; Conservative 59; Mismatches 204;

QY 249 GTTDRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWLGKGLDAAVDDKGT 308
 DB 102 GPTDFKWINALNSCAKVFMAFDSTAPT-----WRNLLA-GQTLVAAV----- 146

QY 309 FLRVLRNDRNTAPGGGFTL---PGRSLMFRVNVG-HLMTNDIAVDTDGSEVEGIMDA 364
 DB 147 -----RGDLSFDAPNGHYALRPEAEAVLIVRPRGHLDKXVLI--DGQFLAGGLFDA 199

QY 365 LFTGLIAHLKASDVNGPLINSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLGLPQN 424
 DB 200 ---ALFAFH-----NGRTLLAKDRGPYLYLPKLSMEEAALNDTALAHTEAMGLPHG 249

QY 425 TWKIGIMDEERTTNLAKICAAADRVVINTGFLDRTGDEIHTSMGAGPMV---RKG 481
 DB 250 QIKVTVLITLPAVPEMDEILYALFERIVGLNCGRWYIFSYLKTFFRAHRDRVLPERGQV 309

QY 482 KSPQWILAYED-----HNVDAAGLAAGSGRAQVKGKGMVMTMELMADVETKIAQPRAG 535
 DB 310 TWQFPLRAYSELLIKTCHRRGAHANGVAAQIPINHDAAEQANMARVADKREVSAQ 369

QY 536 ASTAWVPST-----AATLHALHYQDVAAVQOGLAGKRRATTIEQLITPLA 583
 DB 370 HDGTWVAHPALIPVAMKLFDEHMTAHOQHVLRLNDVQVTR-----DLLIAP-- 415

QY 584 KSLANAPDEI-REEVDNNQCSILGYVVRVVDQGVGSKVPDIHDVVALMEDRATLRISSQL 642
 DB 416 -----SPGNVTRAGFEGNVVCVRYLAALWD-GNGC--VPIHH---LNEADTAETISAQ 464

QY 643 LANWLRHGV-----ITSADVRASLERMAPLVDQVQAGVAVRPMAPNPDSDIAFL 691
 DB 465 LQWQLHHGQHLDDGTAIHQHLLQATLRALPARL-----GTATALPGAARIDEAL 515

QY 692 LAAQELILSGAQO 704
 DB 516 LEE-----LSGAE 524

RESULT 13
 QBPQR6
 ID QBPQR6 PRELIMINARY; PRT; 540 AA.
 AC QBPQR6

Db 609 EKHVLHNQPTSGS-----LFDFGLFVFNHAKA-----LIAQSGSF-YFYLKQLQ 652
QY 401 GPAEVAFTCELFSEVEDVLGLPQNTMKIGIMDEERTTVNLKACIKAAADRVRVINTGFL 460
Db 653 SAEQAQLWADVFKYTKEDKLGARGTIKCTVLIEHLASQLHEIHALKONIVGLNGRW 712
QY 461 DRTGDEIHTSME-----AGPMVRKGTMSQPMILAYEDHNDVAGLAAGFSG 506
Db 713 DYIFSYIKTFQNHKFKLLPDRFQIGMTAPFR-----NYSLEVIKACHLRG 758
QY 507 RAQVCKGMWMTMELMADWV--ETKIAQPR-----GASTAVVPSPTATLHALYHQ 556
Db 759 IHAMG-GWAAQIPKHDQVANDKAFALVRADKEREATDGHGTWVHP-----805
QY 557 VDVAVQOGLAGKRRATIEQLLTP--LAKELAWA-----PDEIREVD--NNCQ 602
Db 806 -----GLVPLAKRVFDQMPKPNQISKNLTRANCTKEDLTVIPEGTRTEAGFRNIS 857
QY 603 SILGYVVRVWDQGVGCSKVPDIHDVLMEDRATLRISQLLANLWRH-----G 650
Db 858 VTGLYDLSWL-RGTGC--VPLYN---LMEDAATAEISRAQLWQNLHHDKLEDRGTIDAG 911
QY 651 VI---TSADVRSASLERMAPLYDR 670
Db 912 LVKQTIAAETERRLLIRAGSVVNR 934

Search completed: November 21, 2003, 16:08:06
Job time : 40.4127 secs

RESULT 15
ID 017353 PRELIMINARY; PRT; 968 AA.
AC 017353;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 108.6 kDa protein.
GN C05E4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleodierinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026209; AAB71278.2; -;
DR HSSP; O53752; 1F61.
DR WormPep; C05E4.9; CE23521.
DR InterPro; IPR006254; Isocit lyase.
DR InterPro; IPR000918; Isocit lyase ph.
DR InterPro; IPR006252; Malate synthase.
DR InterPro; IPR001465; Malate synthase.
DR Pfam; PF00463; ICL; 1.
DR Pfam; PF01274; Malate synthase; 1.
DR ProDom; PD001857; Isocit lyase ph; 1.
DR TIGRFAMs; TIGR01346; isocit lyase; 1.
DR TIGRFAMs; TIGR01344; malate syn A; 1.
DR PROSITE; PS00161; ISOCITRATE_LYASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 968 AA; 108628 MW; 7BAAP979C1F6D31F CRC64;

Query Match 5.2%; Score 198.5; DB 5; Length 968;
Best Local Similarity 23.8%; Pred. No. 0.00044;
Matches 148; Conservative 70; Mismatches 222; Indels 183; Gaps 34;

QY 127 LNAANARNGSLYDALYGTVDVPEPDGAEKGTYNKVRGDKVIAY--ARKELDSVPLSSG 184
Db 416 LSSITALSGSTEEAQFQFAVASQ--DEELSLTAQNAGVAGDEKILTPDALRFLHDL-----458

QY 185 SFGDATGTGVDGQIV-----VALPD--KSTGLANPQFAGYTGAAESPTSV 229
Db 469 ---NTEFNPRRLLSKRNQVQADINNSLWFPDFNKETEVLRSQ--GWKG-AEIPDL 521

QY 230 LLINHLHILIPESQVGTTRAGVKDVLESATITIMDFEDSVAADAAKVLGYRN 289
Db 522 ---QDREVEI-----TGPTDRQWVNAVNSGANVFMAFDENSPST-----WRN 562

QY 290 WL--GLNKGLAAAVDKGTAFRLVNRDRNYTAP--GGQFTLPGR--SLMFRNVG--HLM 344
Db 563 QLEGQINLYD-----AVRNISYTHPTTKKEYTLNEKHAVLKVRPRGWHL 608

QY 345 TNDALV---DTGGEVFEQINDALFT--GLIAHGLKASDVNGPLINSRTGSIYIVKPKMH 400

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 25.6707 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VQYGGSSVADAEIRIRVAE.....SAATRRPCRTGRDGRWACQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 110763 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 110763

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq 19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	DB ID	Description
1	2193	100.0	1172 22 AAU01900	M. tuberculosis Tb
2	1619.5	73.8	421 22 AAU08233	Mycobacterium poly
3	1248.5	56.9	421 22 AAG64046	Corynebacterium th
4	1227.5	56.0	421 21 AAB29607	Modified Corynebac
5	1227.5	56.0	421 24 ABP97764	Amino acid sequenc
6	1224.5	55.8	420 24 ABP97763	Amino acid sequenc
7	1224.5	55.8	421 22 AAU71874	C. glutamicum meta
8	1224.5	55.8	421 22 AAG90027	C. glutamicum prote
9	1224.5	55.8	421 22 AAB79646	Corynebacterium gl

10	1223.5	55.8	421 15 AAR51469	AK alpha subunit T
11	1223.5	55.8	421 15 AAR51465	AK alpha subunit T
12	1223.5	55.8	421 16 AAR74416	C. glutamicum prot
13	1223.5	55.8	421 16 AAR86307	Corynebacterium gl
14	1223.5	55.8	421 19 AAW69147	Aspartokinase alph
15	1223.5	55.8	421 19 AAW69547	Brevibacterium lac
16	1223.5	55.8	421 19 AAW47401	B. lactofermentum
17	1219.5	55.6	421 14 AAR40186	B. flavum AK Bre
18	1219.5	55.6	421 15 AAR51466	AK wild type alpha
19	1219.5	55.6	421 15 AAR51470	AK wild type alpha
20	1219.5	55.6	421 18 AAW06586	B. lactofermentum
21	1219.5	55.6	421 19 AAW69152	Aspartokinase alph
22	1219.5	55.6	842 15 AAR63567	Feedback inhibito
23	1219.5	55.6	842 15 AAR63570	Feedback inhibito
24	1218.5	55.6	421 15 AAR63571	Feedback inhibito
25	1217.5	55.5	421 22 AAE04857	Corynebacterium gl
26	1217.5	55.5	421 23 ABE81982	C. glutamicum aspa
27	1215.5	55.4	421 15 AAR63566	Feedback inhibito
28	1215.5	55.4	842 15 AAR63568	Feedback inhibito
29	1215.5	55.4	842 15 AAR63569	Feedback inhibito
30	1211.5	55.2	421 22 AAE04867	Corynebacterium sp
31	1211.5	55.2	421 23 ABB81989	C. glutamicum aspa
32	1176	53.6	505 22 AAU55464	Propionibacterium
33	787	35.3	409 21 AAB29516	Methylophilus meth
34	785	35.8	411 14 AAR40846	Aspartokinase II.
35	773.5	35.3	254 23 ABP65555	Bifidobacterium lo
36	770	35.1	156 20 AAY39178	M. tuberculosis an
37	770	35.1	156 20 AAY39035	M. tuberculosis re
38	703	32.1	384 24 ABP79567	N. gonorrhoeae ami
39	703	32.1	404 23 ABE48331	Listeria monocytog
40	673	30.7	394 19 AAW98712	H. pylori GHPO 774
41	612.5	27.9	606 11 AAR06915	Protein with aspar
42	563	25.7	424 23 ABP40380	Staphylococcus epi
43	506	23.1	403 23 ABB47953	Listeria monocytog
44	419	19.1	118 23 ABP08315	Human OREF protein
45	377	17.2	564 22 AAE04359	Glycine max aspart

ALIGNMENTS

RESULT 1
AAU01900
ID AAU01900 standard; Protein; 1172 AA.
XX AAU01900;
AC AAU01900;
XX
XX 29-AUG-2001 (first entry)
DT
DE M. tuberculosis TbF14 fusion protein.
XX
XX TbF14; antigen; vaccine; tuberculosis; AIDS; His tag; MtB81; Mo2;
KW acquired immunodeficiency disease.
XX
XX Synthetic.
OS Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT Binding-site 3..8
FT FT /label= Histidine tag
FT FT /note= "Nickel chelating region used for purifying
FT FT the fusion protein"
FT FT 9..749
FT FT /label= MtB81_region
FT FT Misc-difference 49
FT FT /note= "Encoded by TC"
FT FT Region 750..1172
FT FT /label= Mo2_region
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
PD
XX

```

PF 10-OCT-2000; 2000MO-US28095.
XX
XX 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158425.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
XX WPI; 2001-290576/30.
DR N-PSDB; AAS03794.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
XX Claim 3; Fig 3; 168pp; English.
XX
XX The sequence represents Mycobacterium fusion protein antigen TbF14
CC consisting of a His tag for purification, antigen Mb81 and antigen MO2.
CC Compositions comprising at least 2 heterologous antigens, as a fusion
CC protein, and vectors expressing the fusion proteins are used as vaccines
CC to prophylactically immunise mammals (especially humans) against
CC infection by Mycobacteria. The compositions contain at least 2
CC heterologous antigens that increase the serological sensitivity of
CC individuals infected with tuberculosis, a disease frequently affecting
CC patients with acquired immunodeficiency disease, AIDS.
XX
XX SQ Sequence 1172 AA;
XX
XX Query Match 100.0%; Score 2193; DB 22; Length 1172;
XX Best Local Similarity 100.0%; Pred. No. 5.5e-185; Indels 0; Gaps 0;
XX Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VKYGGSSVADAEIRRVAAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPP 60
XX 749 VKYGGSSVADAEIRRVAAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPP 808
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 120
XX 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 868
XX
XX 121 LEEGRVVLVAGFGQVSQDVKDVTTLGRGSGDITAVAMAALGADVCEIYTDVDFGFSADP 180
XX 869 LEEGRVVLVAGFGQVSQDVKDVTTLGRGSGDITAVAMAALGADVCEIYTDVDFGFSADP 928
XX
XX 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
XX 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 988
XX
XX 241 IKDVPMEDEPILGVAHDRSEAKVTIVGLPDIPGYAAKVFRVARRRORHGGAAERLQGR 300
XX 989 IKDVPMEDEPILGVAHDRSEAKVTIVGLPDIPGYAAKVFRVARRRORHGGAAERLQGR 1048
XX
XX 301 GRQDRHLLPQTSGPPPPKNTWTSRSTASTQLLYDDHIGKVLISGAGMRSHPGVTAT 360
XX 1049 GRQDRHLLPQTSGPPPPKNTWTSRSTASTQLLYDDHIGKVLISGAGMRSHPGVTAT 1108
XX
XX 361 FCEALAAVGNIELISTSDQSRCCAAATPNWTRPWSRCMKESGSAATRRPCTRGDRGR 420
XX 1109 FCEALAAVGNIELISTSDQSRCCAAATPNWTRPWSRCMKESGSAATRRPCTRGDRGR 1168
XX
XX 421 WACQ 424
XX 1169 WACQ 1172
XX
XX RESULT 2
XX ID AAU08233
XX XX AAU08233 standard; Protein; 421 AA.
XX AC AAU08233;
XX XX
XX DT 17-DEC-2001 (first entry)

```

```

XX XX Mycobacterium polypeptide for MO-2.
XX DE
XX XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX KW immunostimulant; MO-2.
XX OS
XX XX Mycobacterium sp.
XX XX
XX XX Key Location/Qualifiers
XX FT Misc-difference 41
XX FT Misc-difference /note= "Encoded by T"
XX FT Misc-difference 305
XX FT Misc-difference /note= "Encoded by GAC"
XX XX
XX XX WO200162893-A2.
XX XX
XX XX 30-AUG-2001.
XX XX
XX XX 26-FEB-2001; 2001MO-US05992.
XX XX
XX XX 25-FEB-2000; 2000US-0185037.
XX PR 08-AUG-2000; 2000US-0223828.
XX XX
XX XX (CORI-) CORIXA CORP.
XX XX
XX XX Campos-Neto A, Skeiky Y, Owendale P, Jen S, Lodes M;
XX PI
XX XX WPI; 2001-536638/59.
XX DR N-PSDB; AAS12505.
XX XX
XX XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX XX response to and inhibiting development of a Mycobacterium infection -
XX XX
XX XX Claim 1; Fig 17; 161pp; English.
XX XX
XX XX The present invention relates to the isolation of Mycobacterium
XX CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX CC encoding them. The invention describes compounds and methods for the
XX CC diagnosis of tuberculosis or for inducing protective immunity against
XX CC tuberculosis. The compounds comprise at least one immunogenic portion
XX CC of one or more Mycobacterium proteins and nucleic acid molecules
XX CC encoding such polypeptides. The Mycobacterium proteins and nucleic acid
XX CC molecules encoding them can be used in diagnostic kits for the detection
XX CC of Mycobacterium infection in patients and biological samples. The
XX CC compounds of the invention and antibodies directed against the
XX CC Mycobacterium proteins may be used in vaccines for immunisation against
XX CC Mycobacterium infections. The nucleic acids encoding the Mycobacterium
XX CC proteins may be used in gene therapy. The present sequence represents
XX CC Mycobacterium polypeptide for MO-2.
XX XX
XX SQ Sequence 421 AA;
XX
XX Query Match 73.8%; Score 1619.5; DB 22; Length 421;
XX Best Local Similarity 85.1%; Pred. No. 3.8e-142;
XX Matches 336; Conservative 8; Mismatches 34; Indels 17; Gaps 3;
XX
XX 1 VKYGGSSVADAEIRRVAAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPP 60
XX 5 VKYGGSSVADAEIRRVAAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPP 64
XX
XX 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 120
XX 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 124
XX
XX 121 LEEGRVVLVAGFGQVSQDVKDVTTLGRGSGDITAVAMAALGADVCEIYTDVDFGFSADP 180
XX 125 LEEGRVVLVAGFGQVSQDVKDVTTLGRGSGDITAVAMAALGADVCEIYTDVDFGFSADP 184
XX
XX 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
XX 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 244
XX

```


Query Match 56.0%; Score 1227.5; DB 21; Length 421;
 Best Local Similarity 64.0%; Pred. No. 1.3e-105;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPP 60
 DB 5 VKYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNARIVDVTPGRVREA 124

QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
 DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSRPGTVVVG 240
 DB 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEYARAFNPVLRVRSYSNDPGLIAGS 244

QY 241 IKDVPMEDEPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAHLTGATDKSEAKVTIVGLISDKPEAAKVFRAADAENIDMVLQNVSSVE 304

QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPMKNWTRSETRSASTQLLYDDHIGKVSII 347
 DB 305 DGTDTDIIFTCPRSDGRAMEILKQLVQVQ-----TNVLYDDQVGVKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5

ABP97764
 ID ABP97764 standard; Protein; 421 AA.
 AC ABP97764;
 XX
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Amino acid sequence of an aspartate kinase carrying T3111 mutation.
 XX
 KW Coryneform bacteria; chemical compound; L-amino acid; vitamin; lysC;
 XX aspartate kinase.
 XX
 OS Synthetic.
 OS Corynebacterium glutamicum.
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Met encoded by GTG"
 FT FT
 FT Misc-difference 311 /note= "Met encoded by GTG"
 FT FT
 FT Misc-difference /note= "wild type Thr changed to Ile"
 XX
 PN WO2003014330-A2.
 XX
 XX 20-FEB-2003.
 PD
 XX 30-JUL-2002; 2002WO-EP08465.
 XX
 XX 06-AUG-2001; 2001US-309877P.
 XX
 XX (DEGS) DEGUSSA AG.
 PA
 XX Bathe B, Kreutzer C, Moeckel B, Thierbach G;
 XX
 XX WPI; 2003-278476/27.
 XX
 DR N-PSDB; ABZ77392.
 DR
 XX Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at
 PT least two copies of open reading frame, gene or allele in question at a
 PT particular desired site

XX
 PS Claim 16; Page 105-106; 109pp; English.
 XX
 CC The specification describes Coryneform bacteria which produce chemical
 CC compounds, which instead of a single copy of an open reading frame (ORF),
 CC gene or allele naturally present at a desired locus have at least two
 CC copies of the ORF. Gene or allele, preferably in tandem arrangement.
 CC The modified bacteria are is useful for producing chemical compounds
 CC including L-amino acids, vitamins, nucleosides and nucleosides. The
 CC present sequence is encoded by a lysC gene, and represents an aspartate
 CC kinase carrying the T3111 mutation. Tandem duplications of the lysC
 CC gene are used to construct the bacteria of the invention.

SQ Sequence 421 AA;

Query Match 56.0%; Score 1227.5; DB 24; Length 421;
 Best Local Similarity 64.0%; Pred. No. 1.3e-105;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPP 60
 DB 5 VKYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNAKIIDVTGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAVAIESLGAHARSFTGSOAGVITTCGHNARIVDVTPGRVREA 124

QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGSADP 180
 DB 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSRPGTVVVG 240
 DB 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEYARAFNPVLRVRSYSNDPGLIAGS 244

QY 241 IKDVPMEDEPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAHLTGATDKSEAKVTIVGLISDKPEAAKVFRAADAENIDMVLQNVSSVE 304

QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPMKNWTRSETRSASTQLLYDDHIGKVSII 347
 DB 305 DGTDTDIIFTCPRSDGRAMEILKQLVQVQ-----TNVLYDDQVGVKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 6

ABP97763
 ID ABP97763 standard; Protein; 420 AA.
 AC ABP97763;
 XX
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Amino acid sequence of a wild-type aspartate kinase.
 XX
 KW Coryneform bacteria; chemical compound; L-amino acid; vitamin; lysC;
 XX aspartate kinase.
 XX
 OS Corynebacterium glutamicum.
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Met encoded by ATG"
 FT FT
 FT WO2003014330-A2.
 XX
 XX 20-FEB-2003.
 PD
 XX 30-JUL-2002; 2002WO-EP08465.
 XX
 XX Particular desired site

```

06-AUG-2001; 2001US-309877P.
(DEGS ) DEGUSSA AG.
Bathe B, Kreutzer C, Moeckel B, Thierbach G;
PI
WPI; 2003-278476/27.
N-PSDB; ABZ77391.
Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at
least two copies of open reading frame, gene or allele in question at a
particular desired site
XX
PS Claim 15; Page 102-103; 109pp; English.
XX
CC The specification describes Coryneform bacteria which produce chemical
CC compounds, which instead of a single copy of an open reading frame (ORF),
CC gene or allele naturally present at a desired locus have at least two
CC copies of the ORF, gene or allele, preferably in tandem arrangement.
CC The modified bacteria are is useful for producing chemical compounds
CC including L-amino acids, vitamins, nucleosides and nucleotides. The
CC present sequence is encoded by a wild type lysC gene, and represents
CC aspartate kinase. tandem duplications of the lysC gene are used to
CC construct the bacteria of the invention.
XX
SQ Sequence 420 AA;
Query Match 55.8%; Score 1224.5; DB 24; Length 420;
Best Local Similarity 64.0%; Pred. No. 2.5e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAERIRRVABERIVATKKQGNDDVVVVSAMGDTTDDLLELAQVCPAPP 60
DB 4 VQYGGSSLESABERIRNVAERIVATKKAGNDVVVVC SAMGDTTDELELAANVPVPPAR 63
QY 61 ELDMLLTAGERISNALVMAIESLGAHARFTGSGOAGVITGTGHNKAKI.DVTPERLQTA 120
DB 64 EMDMLLTAGERISNALVMAIESLGAHARFTGSGOAGVITGTGHNARIVDVTPGRVREA 123
QY 121 LEBGRVVLVAGFQGVSDTKDVTTLGRGGSDTTAVAAALGADVCIYTDVDCIFPSADP 180
DB 124 LDEGKICIVAGFCGVNKETDVTTLGRGGSDTTAVALAALNADVCIYSDVDGVYTADP 183
QY 181 RIVRNARLDTVTPEEMLEMAACGAKVLMRCVEYARHNI.PHVRSSYSYSDRPTVVVGS 240
DB 184 RIVPNAQLEKLSPEEMLEAAVGSKILVRSVEYARAFNVPLRVRSYSYSDNPGTLAGS 243
QY 241 IKQVPWEDPILTVGAHDRSEAKVTIVGLPIDIPGYAAKVRAVA-----RRRQHR 290
DB 244 MEDIPVEAVLTGVATDKSEAKVTVLGIDSKPGGAAKVFRALADAEINIMVLGNQVSVE 303
QY 291 HGAERLQGRGRQD-RHLLHLP--QTSGPPPWKNWTRSETRSASTOLLYDDHTGKVS LI 347
DB 304 DGTDTTIFTCPSDGRRAEILKQLVQG-----NW-----TNVLYDDQVGKVS LV 349
QY 348 GAGVRSHPGVTATFCEALAAVGNIELISTSE 379
DB 350 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 381
RESULT 7
AAU71874
ID AAU71874 standard; Protein; 421 AA.
XX
XX AAU71874;
AC
AC
XX
XX
XX 26-FEB-2002 (first entry)
XX
DE C. glutamicum metabolic pathway protein encoded by gene #9.
XX
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
KW methionine biosynthesis pathway; large-scale production of fine chemical;
KW Corynebacterium diphtheriae; diphtheria.

```

```

RESULT 8
AAG90027
ID AAG90027 standard; Protein; 421 AA.
XX AC AAG90027;
XX DT 26-SEP-2001 (first entry)
XX DE C glutamicum protein fragment SEQ ID NO: 3781.
XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis.
XX OS Corynebacterium glutamicum.
XX PN EP1108790-A2.
XX PD 20-JUN-2001.
XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX DR WPI; 2001-376931/40.
XX DR N-PSDB; AAH65246.
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS Claim 17; SEQ ID NO: 3781; 246pp + Sequence Listing; English.
XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived
XX CC from Coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a protein described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.
XX SQ Sequence 421 AA;
Query Match 55.8%; Score 1224.5; DB 22; Length 421;
Best Local Similarity 64.0%; Pred. No. 2.5e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VKYGGSSVADAEIRRVAAERIVATKQGNDDVVVVVVSAMGDTTDLDLAQVCPAPPPR 60
DB 5 VKYGGSSLESARIRNVAERIVATKAGNDVVVVCAMGDTTDLLELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAISSLGNHARSFTSGAGVITGTHGNAKIIVTPCRLQTA 120
DB 65 EMDMLLTAGERISNALVAMAISSLGAEASQFTSGAGVLTTERHGNARIVDTPGVREA 124
QY 121 LEEGRVVLVAGFGQYSQDTKYDTTLTGRGSDTTAVAMAAALGADYCEIYTDVDFGSADP 180
DB 125 LDEGKICIVAGFGQYKNETRDVTLTGRGSDTTAVALAAALNADYCEIYSDVDGVYTDAP 184
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLCQVEVARHNIPVHVRSSYSDRPGTVVVG 240

```

Db 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEYARAFNVPVRVSSVNDGCTLIAGS 244
 QY 241 IKDVPMDPILTGVAHDRSEAKVTIVGLPDPYGAAYKVFRAVA-----RRRQRH 290
 Db 245 MEDIPVEEAVLTGATDKSEAKVTILGISDKPGEAAKVFRAADABINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHLLHLP--QTSGPPPKNKNWTRSETRSASTQLLYDDHIGKVS 347
 Db 305 DGTDTTFTCPRSDDRAMEILKLVQGG-----TNVYDDQVGKVS 350
 QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 351 GAGWKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 9
 AAB79646
 ID AAB79646 standard; Protein; 421 AA.
 XX AC AAB79646;
 XX DT 30-APR-2001 (first entry)
 XX DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:26.
 XX KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
 XX OS Corynebacterium glutamicum.
 XX PN WO200100843-A2.
 XX PD 04-JAN-2001.
 XX PF 23-JUN-2000; 2000WO-IB00923.
 XX PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.

PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-137957/14.
DR N-PSDB; AAF71765.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 20; Page 193-194; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 421 AA;
Query Match 55.8%; Score 1224.5; DB 22; Length 421;
Best Local Similarity 64.0%; Pred. No. 2.5e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNDDVVVVVSAAGDTTDDLLDAQVCPAPPPR 60
Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVCSAGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAESAQSFSGQAGVLTTERHGNARIVDVTGVRREA 124
QY 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDGIFSDP 180
Db 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVYADP 184
QY 181 RIVRNARKLDTVTPEMLEMAACGAKVLMRLCIVEYARRHNPVHVRSSYSYDRPGTVVVG 240
Db 185 RIVPNAQKLEKLSPEMLELAAGVSKILVLSVEYARFNPVPLVRSSYSNDPOTLIAGS 244
QY 241 IKDVPMEPIITGVADHDESEAKVITVGLPDTPGVAANKVRAVA-----RRRQHR 290
Db 245 MEDIPVEEAVLTGVDKSEAKVITVGLISDRKPGGAARVFRALADAEINIDMVLQNVSSVE 304

QY 291 HGAERLQGRQCD-RHHLLLP--QTSPPPKWTRSETRTSASTOLLYDDHIGKVSLLI 347
Db 305 DGTDTITFTCPRSDDRRAWEILKKLQVQG-----TNVLYDDQVGKSLV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
RESULT 10
AARS1469
ID AARS1469 standard; Protein; 421 AA.
XX AARS1469;
XX AC
XX DT 15-NOV-1994 (first entry)
XX DE AK alpha subunit T279A.
XX KW Mutant; C. glutamicum; aspartokinase; AK; alpha; beta; subunit;
KW substitution; non-acidic; amino acid; Ala; additional; enzyme;
KW feedback inhibition; synergic inhibition; heat stability.
XX OS Corynebacterium glutamicum.
XX
FH Key Location/Qualifiers
FT Misc-difference 279 /label= T279A
XX
PN JP06062866-A.
XX
PD 08-MAR-1994.
XX
PF 27-APR-1993; 93JP-0101450.
XX
PR 28-APR-1992; 92JP-0110292.
XX
PA (AJIN) AJINOMOTO KK.
XX
XX WPI; 1994-121127/15.
DR N-PSDB; AAQ61595.
XX
XX The mutant aspartokinase gene - used to produce AK which is
PT removed from feedback inhibition
XX
PS Disclosure; Page 14-16; 28pp; Japanese.
XX
XX This sequence represents a mutant version of the C. glutamicum asparto-
CC kinase (AK) alpha subunit. The mutation is at position 279 and is a
CC corresponds to the substitution of a non-acidic amino acid, pref.
CC Ala. This mutation causes removal of the enzyme from additional
CC feedback inhibition. Inhibition by Thr is removed completely and
CC the synergic inhibition of Lys and Thr is also removed. The new AK
CC has increased heat stability.
XX
SQ Sequence 421 AA;
Query Match 55.8%; Score 1223.5; DB 15; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.1e-105;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNDDVVVVVSAAGDTTDDLLDAQVCPAPPPR 60
Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVVCSAGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAESAQSFSGQAGVLTTERHGNARIVDVTGVRREA 124
QY 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDGIFSDP 180
Db 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVYADP 184

QY 181 RIVNARKLDTVTFEEMLEWAGCAKVMRLRCVEYARRHNPVHVRSSYSYSDRPGTVVGS 240
 DB 185 RIVPNAQKLEKSFEMLELAAVGSKILVRSVEYARFNPVLRVSSYNDPGLIAGS 244
 QY 241 IKDVPMDPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAVLTGVATDKSEAKVTILGISDKPGEAAKVFRAALADAEINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHHHLHP--QTSGPPPPKWNTRSETRSASTQLLYDDHIGKVSLL 347
 DB 305 DGTDTITFTCPRADGRAMEILKLVQVQ-----TNVLYDDQVGKVSLLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 11

AAR51465
 ID AAR51465 standard; Protein; 421 AA.

AC AAR51465;

DT 15-NOV-1994 (first entry)

XX AK alpha subunit T279A from clone ATCC 13869.

XX Mutant; C. glutamicum; aspartokinase; AK; alpha; beta; subunit;
 KW substitution; non-acidic; amino acid; Ala; additional; enzyme;
 KW feedback inhibition; synergic inhibition; heat stability.

XX Corynebacterium glutamicum.

XX Key Location/Qualifiers
 FT Misc-difference 279 /label= T279A

XX JP06062866-A.
 XX 08-MAR-1994.

XX 27-APR-1993; 93JP-0101450.

XX 28-APR-1992; 92JP-0110292.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1994-121127/15.

XX N-PSDB; AAQ61599.

XX The mutant aspartokinase gene - used to produce AK which is
 FT removed from feedback inhibition

XX Disclosure; Page 22; 28pp; Japanese.

XX This sequence represents a mutant version of the C. glutamicum asparto-
 CC kinase (AK) alpha subunit. The mutation is at position 279 and
 CC corresponds to the substitution of a non-acidic amino acid, pref.
 CC Ala. This mutation causes removal of the enzyme from additional
 CC feedback inhibition. Inhibition by Thr is removed completely and
 CC the synergic inhibition of lys and Thr is also removed. The new AK
 CC has increased heat stability.

XX Sequence 421 AA;

XX Query Match 55.8%; Score 1223.5; DB 15; Length 421;
 XX Best Local Similarity 64.0%; Pred. No. 3.le-105;
 XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDLDDLAQVCPAPPPR 60
 DB 5 VKYGGSSVADAEIRIRVAERIVATKKACNDVVVVSAMGDTTDLLELAAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITGTGHNAKIIDVTPGRLQTA 120
 DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITGTGHNARIVDVTPGVRQA 124
 QY 121 LEBGRVVLVAGFOGVGSOOTKDVTTTLGRGSDTTAVAVAAALGADVCEIYTDVDGIFSADP 180
 DB 125 LDEKICIVAGFOGVGKTRDVTTLGRGSDTTAVAVAAALGADVCEIYSDVDGVYTADP 184
 QY 181 RIVNARKLDTVTFEEMLEWAGCAKVMRLRCVEYARRHNPVHVRSSYSYSDRPGTVVGS 240
 DB 185 RIVPNAQKLEKSFEMLELAAVGSKILVRSVEYARFNPVLRVSSYNDPGLIAGS 244
 QY 241 IKDVPMDPILTCVAHDRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 290
 DB 245 MEDIPVEEAVLTGVATDKSEAKVTILGISDKPGEAAKVFRAALADAEINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHHHLHP--QTSGPPPPKWNTRSETRSASTQLLYDDHIGKVSLL 347
 DB 305 DGTDTITFTCPRADGRAMEILKLVQVQ-----TNVLYDDQVGKVSLLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 12

AAR71416
 ID AAR71416 standard; Protein; 421 AA.

AC AAR71416;

DT 18-JAN-1996 (first entry)

XX C. glutamicum protein.

XX Phosphoenolpyruvate-carboxylase; Corynebacterium glutamicum.

XX Corynebacterium glutamicum.

XX WO9506114-A1.

XX 02-MAR-1995.

XX 17-AUG-1994; 94WO-JP01365.

XX 24-AUG-1993; 93JP-0209775.

XX 24-AUG-1993; 93JP-0209776.

XX 05-JUL-1994; 94JP-0153876.

XX (AJIN) AJINOMOTO KK.

XX Izui K, Matsui H, Sugimoto M, Suzuki T;

XX WPI; 1995-106843/14.

XX N-PSDB; AAQ84970.

XX Variant of phospho-enol pyruvate carboxylase - not substantially
 PT inhibited by aspartic acid, is used for efficient production of
 PT amino acids

XX Disclosure; Page 48-51; 77pp; Japanese.

XX The primers given in AAQ84968-69 were used in the PCR amplification
 CC of C. glutamicum ATCC 13869 DNA to isolate the DNA sequences given
 CC in AAQ84970-71.

XX Sequence 421 AA;

XX Query Match 55.8%; Score 1223.5; DB 16; Length 421;
 XX Best Local Similarity 64.0%; Pred. No. 3.le-105;
 XX Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

PT - used in preparation of L-lysine
 PS Claim 2; Page 17-18; 50pp; English.
 XX
 XX This is the wild-type alpha subunit of Brevibacterium lactofermentum
 CC ATCC 13689 (Corynebacterium glutamicum) aspartokinase, an enzyme
 CC involved in the biosynthesis of L-lysine, but which is subject to
 CC feedback inhibition. A claimed recombinant DNA molecule that is
 CC autonomously replicable in coryneform bacteria comprises a DNA
 CC sequence coding for an aspartokinase in which feedback inhibition
 CC by L-lysine and L-threonine is desensitized, and a DNA sequence
 CC coding for a diaminopimelate decarboxylase. This mutant
 CC aspartokinase preferably has residue Ala-279 of the alpha subunit
 CC (see AAW68152) and residue Ala-30 of the beta subunit (see AAW68153)
 CC replaced by an amino acid that is not Ala and is not an acidic
 CC amino acid. The host coryneform bacteria are used in a claimed
 CC method for producing L-lysine, which is useful as a fodder
 CC additive. They show improvements in L-lysine yield without
 CC suffering restraint of growth.
 XX
 SQ Sequence 421 AA;
 CC
 CC Query Match 55.8%; Score 1223.5; DB 19; Length 421;
 CC Best Local Similarity 64.0%; Pred. No. 3.le-105;
 CC Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRRVAEIRVATKQGNVWVVSAMGDTTDLDDLAQVCPAPP 60
 Db 5 VKYGGSSLESABIRINVAERIVATKAGNDVVVCSAMGDTTDLLELAANVPVPP 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 Db 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNARIVDVTPGRVREA 124
 QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSGDTTAVAMAALGADYCEIYTDVGGIFSADP 180
 Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVALAALNADYCEIYSDVGGVITADP 184
 QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLCVEYARRHNIPVHVRSSYSDRPGTVVVG 240
 Db 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLRVSRSSYNDPGTLLIAGS 244
 QY 241 IKDVPMEDELITGVADHRSKAVTIYGLPDIGYAAKVFRVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGVATDKSEAKVTVLGISDKRGEAAKVFRLADAEINIDVQLQNVSSVE 304
 QY 291 HGAERLQGRGRQD-RHHLHLLP--QTSGPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 347
 Db 305 DGTDTITFCPRADGRAMEILKQLVQG-----NW-----TNVLYDDQVGKVS LV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
 RESULT 15
 AAW69547
 ID AAW69547 standard; Protein; 421 AA.
 XX
 AC AAW69547;
 XX
 XX 13-OCT-1998 (first entry)
 DT
 XX Brevibacterium lactofermentum lysC protein.
 DE
 XX Brevibacterium lactofermentum; lysC; L-lysine; coryneform bacterium;
 KW aspartokinase; feedback inhibition; dihydrodipicolinate reductase;
 KW diaminopimelate decarboxylase; aspartate aminotransferase.
 XX
 OS Brevibacterium lactofermentum.
 XX
 PN EP854189-A2.
 XX

PD 22-JUL-1998.
 XX
 PF 05-DEC-1997; 97EP-0121443.
 XX
 PR 05-DEC-1996; 96JP-0325659.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI Araki M, Nakamatsu T, Sugimoto M, Yoshihara Y;
 XX WPI; 1998-379060/33.
 DR N-PSDB; AAV40254.
 XX
 PT Recombinant DNA autonomously replicable in coryneform bacteria -
 PT used to produce L-lysine, codes for e.g. aspartokinase,
 PT dihydrodipicolinate reductase and synthase and di-amino-pimelate
 PT decarboxylase
 XX
 PS Claim 2; Page 22-23; 59pp; English.
 XX
 CC The present invention describes a recombinant DNA autonomously
 CC replicable in cells of coryneform bacteria (CB), comprising a DNA
 CC sequence coding for an aspartokinase (AK) in which feedback inhibition
 CC by L-lysine and L-threonine is desensitized, a DNA sequence coding for
 CC a dihydrodipicolinate reductase (DHP), a DNA sequence coding for
 CC dihydrodipicolinate synthase (DHPs), a DNA sequence coding for
 CC diaminopimelate decarboxylase (DAMD) and a DNA sequence coding for
 CC aspartate aminotransferase (AAT). The present sequence is wild-type
 CC lysC from Brevibacterium lactofermentum. The DNA and related products
 CC from the present invention, can be used for improving L-lysine
 CC productivity by CB. The L-lysine produced can be used as a fodder
 CC additive.
 XX
 SQ Sequence 421 AA;
 CC
 CC Query Match 55.8%; Score 1223.5; DB 19; Length 421;
 CC Best Local Similarity 64.0%; Pred. No. 3.le-105;
 CC Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VKYGGSSVADAEIRRVAEIRVATKQGNVWVVSAMGDTTDLDDLAQVCPAPP 60
 Db 5 VKYGGSSLESABIRINVAERIVATKAGNDVVVCSAMGDTTDLLELAANVPVPP 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 Db 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTHGNARIVDVTPGRVREA 124
 QY 121 LEEGRVVLVAGFGVSGDVKDTTLGRGSGDTTAVAMAALGADYCEIYTDVGGIFSADP 180
 Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDTTAVALAALNADYCEIYSDVGGVITADP 184
 QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLCVEYARRHNIPVHVRSSYSDRPGTVVVG 240
 Db 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLRVSRSSYNDPGTLLIAGS 244
 QY 241 IKDVPMEDELITGVADHRSKAVTIYGLPDIGYAAKVFRVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGVATDKSEAKVTVLGISDKRGEAAKVFRLADAEINIDVQLQNVSSVE 304
 QY 291 HGAERLQGRGRQD-RHHLHLLP--QTSGPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 347
 Db 305 DGTDTITFCPRADGRAMEILKQLVQG-----NW-----TNVLYDDQVGKVS LV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382
 Search completed: November 21, 2003, 16:03:11
 Job time : 27.6707 secs

Db 5 VQYGGSSVADADRIRRAERIVQTKQGNDIVVWVSAMGDTTDDLLDLAQVCPPEPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124
QY 121 LBEGRVVLVAGFGVQSQDTKDVTTILGRGSDTTAVAMAAALGADVCEIYTDVVDGIFSDAP 180
Db 125 LDEGRVVLVAGFGVQSQDTKDVTTILGRGSDTTAVAMAAALGADVCEIYTDVVDGIFSDAP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 244
QY 241 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRVAAVRRRQRRHGAEE----- 295
Db 245 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRVAAVRRRQRRHGAEE----- 295
QY 296 RIQGRGRDRHHLH---LPQTSPPPKWTRSETRSASTOLLYDDHIGKVSIGAMR 352
Db 296 VLQNVSKVEDGKTDITFTCRSDSGPIAVAKGLSLRDEIGFTQLLYDDHIGKVSIGAMR 355
QY 353 SHPGVTATFCEALAAVGVNIELISTSEDOQRSCCAAT 389
Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 2

US-08-532-828B-3
; Sequence 3, Application US/08532828B
; Patent No. 5688671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,828B
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-101450
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-764-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13869
US-08-532-828B-3
Query Match 55.8%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADADRIRRAERIVQTKQGNDIVVWVSAMGDTTDDLLDLAQVCPPEPPAR 60
Db 5 VQYGGSSLESABRIIRNVAERIVATKAGNDVVVVCAMGDTTDELELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNARIVDVTPGRVREA 124
QY 121 LBEGRVVLVAGFGVQSQDTKDVTTILGRGSDTTAVAMAAALGADVCEIYTDVVDGIFSDAP 180
Db 125 LDEGKICIVAGFGVQNKTRDVTTLGRGSDTTAVAAAAALNADVCEIYSDVDGVYADP 184
QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVLRSEVARENVPLRVSRSSISNDPGTLIAGS 244
QY 241 IKDVPMEDEPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRVAAVRRRQRRHGAEE----- 290
Db 245 MEDIPVEEAVLTGVTADKSEAKVTIVGLISDKPCEAAKVPALADAEINIDMVLQNVSSVE 304
QY 291 HGAERLQGRGRD-RHHLHLLP--QTSPPPKWTRSETRSASTOLLYDDHIGKVSIL 347
Db 305 DGTDTITFTCPRADGRRAEMEILKLVQVQ-----NW-----TNVLYDDQVQKVSILV 350
QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
Db 351 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 382

RESULT 3

US-08-700-359-9
; Sequence 9, Application US/08700359
; Patent No. 5766925
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, MASAKAZU
; APPLICANT: USUDA, YOSHIHIRO
; APPLICANT: SUZUKI, TOMOKO
; APPLICANT: TANAKA, AKIKO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,359
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-35019
; FILING DATE: 04-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618

```
; REFERENCE/DOCKET NUMBER: 10-819-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-700-359-9

Query Match 55.8%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.le-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VKYGGSSLESASERIRVAERIVATKKAGNDVVVCSAMGDTTDELLEAAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIVDTPGRLQTA 120
Db 65 EMDMLLTAGERISNALVAMAIESLGAQAQSFSGQAGVLTTERHGNARIVDVTGVRREA 124
QY 121 LEEGRVVLVAGFQVQSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGFSADP 180
Db 125 LDEGKICIVAGFQVGNKTRDVTTLGRGSDTTAVALAALNADVCIEIYSDVDGVYVADP 184
QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRGTVVVG 240
Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSYARAFNVPURVRSYSDNPGTLLIAGS 244
QY 241 IKDVPMDPILTVGAHDSRAKVTIVGLPIGPAKVFRAVA-----RRRRQHR 290
Db 245 MEDIPVEEAVLTGATDKSEAKTVLGISDKPGEAKVFRALADAEINIDMWLQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSETRSASTQLLYDDHIGKVSLLI 347
Db 305 DGTDTITFTCPRADGRAMEILKQLVQG-----TNVLYDDQVGKVSIV 350
QY 348 GAGMRSHPGVTATFCEALAAVGNVIELISTSE 379
Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 4
US-08-596-366-6
; Sequence 6, Application US/08596366
; Patent No. 5876983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE.
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596.366
; FILING DATE: 29-APR-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-784-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-596-366-6

Query Match 55.8%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.le-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VKYGGSSLESASERIRVAERIVATKKAGNDVVVCSAMGDTTDELLEAAAVNPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIVDTPGRLQTA 120
Db 65 EMDMLLTAGERISNALVAMAIESLGAQAQSFSGQAGVLTTERHGNARIVDVTGVRREA 124
QY 121 LEEGRVVLVAGFQVQSQDTKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGFSADP 180
Db 125 LDEGKICIVAGFQVGNKTRDVTTLGRGSDTTAVALAALNADVCIEIYSDVDGVYVADP 184
QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRGTVVVG 240
Db 185 RIVPNAQKLEKLSFEEMLELAAGSKILVRSYARAFNVPURVRSYSDNPGTLLIAGS 244
QY 241 IKDVPMDPILTVGAHDSRAKVTIVGLPIGPAKVFRAVA-----RRRRQHR 290
Db 245 MEDIPVEEAVLTGATDKSEAKTVLGISDKPGEAKVFRALADAEINIDMWLQNVSSVE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSETRSASTQLLYDDHIGKVSLLI 347
Db 305 DGTDTITFTCPRADGRAMEILKQLVQG-----TNVLYDDQVGKVSIV 350
QY 348 GAGMRSHPGVTATFCEALAAVGNVIELISTSE 379
Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5
US-08-967-104-6
; Sequence 6, Application US/08967104
; Patent No. 5919694
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHOENOLPYRUVATE CARBOXYLASE.
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
```

ADDRESSEE: P.C.
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,104
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,366
FILING DATE: 29-APR-1996
APPLICATION NUMBER: JP 5-203775
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-203776
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-153876
FILING DATE: 05-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-104-6

Query Match 55.8%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPR 60
DB 5 VQYGGSSLESABIRIRNVAERIVATKKAGNDVWVVSAMGDTTDELELAANVPVPPAR 64
QY 61 ELDMMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTGRLQTA 120
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNARIVDVTGVRREA 124
QY 121 LSEGRVVLVAGFGVQSDTKVDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFPSADP 180
DB 125 LDEGKICIVAGFGVQGNKTRDVTTLGRGSDTTAVALLAALNADYCEIYSDVDGVYTDAP 184
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSDRPGTVVVG 240
DB 185 RIVPNQAQLEKLSFEEMLELAAGVSKILVRSVEYARAFNPLRVRSYSDNPGTLIAGS 244
QY 241 IKDVPMDPILTVGAHDSREAKVTIVGLPDIIFYAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGEAQVFRALADAEINIDVQLQNVSSVE 304
QY 291 HGAERLQGRGD-RHLLHLP--QTSGPPPKWNTRETSASSTQLLYDDHIGKVS LI 347
DB 305 DGTDTITFTCFRAGRAMEILKQLVQVQ-----TNVLYDDQVGKVS LI 350
QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 6
US-08-985-908-5
Sequence 5, Application US/08985908
Patent No. 6004773
GENERAL INFORMATION:
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI NA
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,908
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-325659
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-908-5

Query Match 55.8%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQQVCPAPPR 60
DB 5 VQYGGSSLESABIRIRNVAERIVATKKAGNDVWVVSAMGDTTDELELAANVPVPPAR 64
QY 61 ELDMMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTGRLQTA 120
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNARIVDVTGVRREA 124
QY 121 LSEGRVVLVAGFGVQSDTKVDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFPSADP 180
DB 125 LDEGKICIVAGFGVQGNKTRDVTTLGRGSDTTAVALLAALNADYCEIYSDVDGVYTDAP 184
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSDRPGTVVVG 240
DB 185 RIVPNQAQLEKLSFEEMLELAAGVSKILVRSVEYARAFNPLRVRSYSDNPGTLIAGS 244
QY 241 IKDVPMDPILTVGAHDSREAKVTIVGLPDIIFYAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPGEAQVFRALADAEINIDVQLQNVSSVE 304
QY 291 HGAERLQGRGD-RHLLHLP--QTSGPPPKWNTRETSASSTQLLYDDHIGKVS LI 347
DB 305 DGTDTITFTCFRAGRAMEILKQLVQVQ-----TNVLYDDQVGKVS LI 350
QY 348 GAGMRSHPGVTATFCEALAAVGNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

Db 351 GAGKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 8
US-08-985-916-5
; Sequence 5, Application US/08985916
; Patent No. 6221636
; GENERAL INFORMATION:
; APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI
; APPLICANT: MASAKO IGU, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI
; APPLICANT: NAKAMATSU
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; ZIP: 22152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,916
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325658
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-985-916-5

Query Match 55.8%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 1 VKYGGSSVADAEIRIRVAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPPP 60
Db 5 VKYGGSSLESARIRNVAERIVATKKAGNDVVVCSAMGDTTDELELAANVPVPP 64

Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAESAQSFSGQAGVLTTERHGNARIVDTPGRVREA 124

Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVVGIFGADP 180
Db 125 LDEGKICIVAGFGVQGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVPNAQKLEKLSPEEMLELAAGVKILVRSVEYARAFNPLVRSSYNDPGLIAGS 244

Qy 241 IKDVPMDPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRVA-----RRRQHR 290
Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPGEAAKVFRLADAENIDMVLQNVSSVE 304

Qy 291 HGAERLQGRQD--RHHLLIP--QTSPPPPKQWTRSETRSASTOLLYDDHIGKVSLLI 347
Db 305 DGTDTITTCPRADGRRAEILKKLVQVQ-----TNVLYDDQVGVKSLV 350

Qy 348 GAGMRSHPGVTATPCEALAAVGVNIELISTSE 379

US-08-852-730-14
; Sequence 14, Application US/08852730
; Patent No. 6090597
; GENERAL INFORMATION:
; APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EIICHI NAKANO,
; APPLICANT: MASAKO IGU, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI
; APPLICANT: NAKAMATSU
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; ZIP: 22026
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,730
; FILING DATE: 05-07-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-142812
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-852-730-14

Query Match 55.8%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 1.1e-116;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

Qy 1 VKYGGSSVADAEIRIRVAERIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPAPPP 60
Db 5 VKYGGSSLESARIRNVAERIVATKKAGNDVVVCSAMGDTTDELELAANVPVPP 64

Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIVDTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAESAQSFSGQAGVLTTERHGNARIVDTPGRVREA 124

Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVVGIFGADP 180
Db 125 LDEGKICIVAGFGVQGNKTRDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYTADP 184

Qy 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSRDPGTVVVGS 240
Db 185 RIVPNAQKLEKLSPEEMLELAAGVKILVRSVEYARAFNPLVRSSYNDPGLIAGS 244

Qy 241 IKDVPMDPILTVGAHDRSEAKVTIVGLPDIPGYAAKVFRVA-----RRRQHR 290
Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPGEAAKVFRLADAENIDMVLQNVSSVE 304

Qy 291 HGAERLQGRQD--RHHLLIP--QTSPPPPKQWTRSETRSASTOLLYDDHIGKVSLLI 347
Db 305 DGTDTITTCPRADGRRAEILKKLVQVQ-----TNVLYDDQVGVKSLV 350

Qy 348 GAGMRSHPGVTATPCEALAAVGVNIELISTSE 379

Db 351 GAGKSHPGVTAFFMEALRDVAVNIELISTSE 382
 RESULT 9
 US-08-532-828B-4
 ; Sequence 4, Application US/08532828B
 ; Patent No. 568671
 ; GENERAL INFORMATION:
 ; APPLICANT: SUGIMOTO, Masakazu
 ; APPLICANT: OGAWA, Yuri
 ; APPLICANT: SUZUKI, Tomoko
 ; APPLICANT: TANAKA, Akiko
 ; APPLICANT: MATSUI, Hiroshi
 ; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
 ; CITY: ARLINGTON
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MS-DOS TEXT EDITOR
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/532,828B
 ; FILING DATE: 27-OCT-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 5-101450
 ; FILING DATE: 27-APR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: NORMAN F. OBLON
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 10-764-0 PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000
 ; TELEFAX: 703-413-2220
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 421 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Corynebacterium glutamicum
 ; STRAIN: AJ3463
 ; US-08-532-828B-4
 Query Match 55.6%; Score 1219.5; DB 1; Length 421;
 Best Local Similarity 63.8%; Pred. No. 2.8e-116;
 Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPPP 60
 Db 5 VQYGGSSLESABIRIRVAERIVATKKAGNDVVVWCSAMGDTTDELLELAANVPVPPAR 64
 QY 61 ELDMLTAGERISNALVANAIESLGAHARSFTGSGQAGVITTTGHNAKIIDVTPGRLQTA 120
 Db 65 EMDMLTAGERISNALVANAIESLGAHARSFTGSGQAGVITTTGHNARIVDVTGVRREA 124
 QY 121 LEEGRVVLVAGFGVGSQDTKDVTTLGRGSDTTAVAAAAALGADVCEIYTDVDFGSADP 180
 Db 125 LDEGKICIVAGFGVGNKETRDVTTLGRGSDTTAVAAAAALNADVCEIYSDVDGVYADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSRDPGTVVVGS 240

Db 185 RIVNNAQKLEKLSFEEMLELAAGVKSKILVLRSEYARAFNVLPLVRSSYSNDPGTTLIAGS 244
 QY 241 IKQVPMEDPILTCVAHDSREAKVTIVGLPDPICGAAKVPRAVA-----RRRQHR 290
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGISDKPGETAKVFRALADAENIDMVLQNVSSVE 304
 QY 291 HGAERLQGRQD-RHHLHLP--QTSGPPPKWNRSETRSASTQLLYDDHIGKVSII 347
 Db 305 DGTDTITFTCPRADGRAMEILKLVQV-----NW-----INVLYDDQGVKSVL 350
 QY 348 GAGMRSHPGVTATFCALAAVGNIELISTSE 379
 Db 351 GAGKSHPGVTAFFMEALRDVAVNIELISTSE 382
 RESULT 10
 US-09-252-991A-29720
 ; Sequence 29720, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 29720
 ; LENGTH: 461
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-29720
 Query Match 37.5%; Score 823; DB 4; Length 461;
 Best Local Similarity 42.6%; Pred. No. 1.3e-75;
 Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPPP 60
 Db 54 VQKEGTSVCTVERIEQVAEKVKFREAGDDVVVVVSAMSGEINRLIGLANQIMEQVPR 113
 QY 61 ELDMLTAGERISNALVANAIESLGAHARSFTGSGQAGVITTTGHNAKIIDVTPGRLQTA 120
 Db 114 ELDVWVSTGEQVTALLSMALIKRGVPAVSYTGQVRILTDSHTKARILHDDTHRAD 173
 QY 121 LEEGRVVLVAGFGVGSQDTKDVTTLGRGSDTTAVAAAAALGADVCEIYTDVDFGSADP 180
 Db 174 LKAGRVVVVAGFGV-DGNGNITTLGSGSDTTGVALAALKADECQIYTDVDSVTTDP 232
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSRDPGTVVVGS 240
 Db 233 RVVQPARLDKITFEEMLEMASLGSVKVQIYRAVEFAGKYNVPLRVLHSHFQEGFTLITD 292
 QY 241 IKQVPMEDPILTCVAHDSREAKVTIVGLPDPICGAAKVPRAVARRRQ----- 288
 Db 293 DEESMEQPIISGIAFNDRDEAKTIRGVPDTPGVAFKILGPISAANVEVDIMVQNVADN 352
 QY 289 -----HRHGAERLQGRQDRHHLHLLPOTSGPPPKWNRSETRSASTQLLYDDH 340
 Db 353 TDTFTFVHRNDVNALE-----ILKQTA-----ANTIGAREALGDTN 389
 QY 341 IGKVSLLGAGMRSHPGVTATFCALAAVGNIELISTSE 379
 Db 390 IAKVSIIVGVMRSHAGVASRMFEALAKESINIQMISTSE 428
 RESULT 11
 US-07-684-135A-2
 ; Sequence 2, Application US/07684135A

Patent No. 5243039
GENERAL INFORMATION:
APPLICANT: Schendel, Frederick J. and
APPLICANT: Flickinger, M. C.
TITLE OF INVENTION: Bacillus MGA 3 Aspartokinase
TITLE OF INVENTION: II Gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
COMPUTER: No. 5243039thgate 386
OPERATING SYSTEM: DOS 4.0
SOFTWARE: Wordperfect- 5.0
CURRENT APPLICATION DATA: US/07/684,135A
FILING DATE: 19910412
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 600,216-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 332-5300
TELEFAX: (612) 332-9081
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
FEATURE:
NAME/KEY: Aspartokinase II ~ dimer subunit
LOCATION: 1 to 411
US-07-684-135A-2

Query Match 35.8%; Score 785; DB 1; Length 411;
Best Local Similarity 42.1%; Pred. No. 8.4e-72;
Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;
QY 1 VKYGGSSVADAERIRRAERIVATKQGNQNDVVVWSAMGDTTDDLLDQAQVCPAPPPR 60
DB 5 VKYGGSSVADAERIRRAERIVATKQGNQNDVVVWSAMGDTTDDLLDQAQVCPAPPPR 64
QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTHGNNAKIIDVTGRLQTA 120
DB 65 EMDMLTTGEQVITSLAMALNEKGYEAI SYTGQAGITTEPVFGNARILNIETEKIQKQ 124
QY 121 LEEGRVVLVAGFQGVSDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVGDGIFADP 180
DB 125 LNEGRVVLVAGFQGVSDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVGDGIFADP 183
QY 181 RIVNRARKLDTVTPEEMLEMAACGAKVLMRCVEYAREHNPVHVRSSYSRDPCTGVVGS 240
DB 184 RYVKSARKLASIDYEMLEMLNGLGVLPRAVEFAKNGITLVRSMSEREETIIE-- 241
QY 241 IKDVEPE-DPILTVGAHDSREAKVTIVGLPIPGYAAKVFRAVARRRRQRHGAERLQG 299
DB 242 -EETVMEQNLVVRGVAEDETITRVTFGLPNLSLSTIFTLA----- 284
QY 300 RGRDRPHLLHPQSGPPPKWKNRSTRSASTQLLYDD----- 339
DB 285 ---QNRINVDIIQSA-----TDAETNLSFIKSDSDLEETVAVLENNKLLNYQGI 333
QY 340 ---HIGKVLISAGMRGHPGVTAFCALAAVGNIELISTSE 379
DB 334 ESETGLAKVISVGSWMINPGVAQWFEVLALNGIQWQVSTSE 377

RESULT 12

US-09-056-556-227
Sequence 227, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 227:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-056-556-227

Query Match 35.1%; Score 770; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.5e-71;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 QKYGGSSVADAERIRRAERIVATKQGNQNDVVVWSAMGDTTDDLLDQAQVCPAPPPR 61
DB 1 QKYGGSSVADAERIRRAERIVATKQGNQNDVVVWSAMGDTTDDLLDQAQVCPAPPPR 60
QY 62 LDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTHGNNAKIIDVTGRLQTA 121
DB 61 LDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTHGNNAKIIDVTGRLQTA 120
QY 122 EGRVVLVAGFQGVSDTKVDVTTLGRGSDTTAVAM 157
DB 121 EGRVVLVAGFQGVSDTKVDVTTLGRGSDTTAVAM 156

RESULT 13

US-09-072-596-222
Sequence 222, Application US/09072596
Patent No. 6459366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 222:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 156 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-072-596-222

Query Match 35.1%; Score 770; DB 4; Length 156;
 Best Local Similarity 100.0%; Pred. No. 6.5e-71;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 QKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPRE 61
 DB 1 QKYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPRE 60
 QY 62 LDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 121
 DB 61 LDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 120
 QY 122 EGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAM 157
 DB 121 EGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAM 156

RESULT 14
 US-09-328-352-6511
 ; Sequence 6511, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCES: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6511
 ; LENGTH: 438
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-6511

Query Match 34.0%; Score 746.5; DB 4; Length 438;
 Best Local Similarity 42.0%; Pred. No. 8.2e-68;
 Matches 168; Conservative 73; Mismatches 124; Indels 35; Gaps 8;

QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 DB 17 VQYGGTSMGTTPERILNVAERVKRWHDHGKVVVVVSAMSGETNRLLAKALETDPDPR 76
 QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 120
 DB 77 ELDQMVSTGEQVTISMLAVALSIGVEAKSYTGROVGKTDSTAKTARIESIDTVMTND 136
 QY 121 LEBGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 180
 DB 137 LDAGRIVVAGFQGFQDAN-GNTTTLGRGSGDTSGVAIAAALKADECIYTDVGVVYTDTP 195
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMURCQVEYARRHNPVHVRSS----- 228
 DB 196 RVAPKAKKIDRISFEEMLEMASLGSKVLQIRSVFPAKYQVPLRVLSPPDNDGDDGAFDDE 255
 QY 229 YSDRPGTVVVGSIKDVEMEDPILTVGAHDSRAKVTIVGLPDIPGYAAKVRAVAR--- 285
 DB 256 FKENVGTLLITAEADT-MEQPIIAGIAFNRDEAKLITLGVDPFGIASKILSPVSDANIE 314
 QY 286 -----RQHRHGAERLQGRGDRHLLH-LLPQTSPPPKWNTSRSETASATQLLYDD 339
 DB 315 VDMIVQNVBEDGTTDFTTVNRVDLAKAEKILNETA-----KN---IGAREVSTR-----D 362
 QY 340 HIGKVSILGAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 363 DIVKVSIVGVGMRSHAGVASKMFTALADEGINILMISTSE 402

RESULT 15
 US-09-134-001C-5225
 ; Sequence 5225, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5225
 ; LENGTH: 424
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-5225

Query Match 25.7%; Score 563; DB 4; Length 424;
 Best Local Similarity 35.6%; Pred. No. 4.8e-49;
 Matches 138; Conservative 82; Mismatches 142; Indels 26; Gaps 8;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 DB 29 VLKPGSSVSDFKKIRNIAEMKLTREDGEEILVVVSAMGKTTDQLMNVSSLTSTPKQ 88
 QY 61 ELDMLLTAGERISNALVAMAEISGAHARSFTGSQAGVITGTHGNAKIIDVTPGRLQTA 120
 DB 89 ELALLTTGEGQVSVLSVNLNDIGVNAKAMTGYAGIKTVGHHLASKIAENPNIFNEA 148
 QY 121 LEBGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 180
 DB 149 FKNDILVAGFQGINEDF-ELTTLGRGSGDTTAVALAAS-NQTPCEIYTDVGVYATDP 206
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMURCQVEYARRHNPVHVRSSYSDRPGTVVVG 240
 DB 207 RIHNEAKLEYSVEENWMSALGAGVLETRVELAKNYDIPLYLGTLSNVKGTWIMS- 265
 QY 241 IKQVPMEDPILTVGAHDSRAKVTI-VGLDPIPGYAAKVRAVARRRRRHGAERLQ 299

```
8
Db 266 -KSDLEKKAVTGVALDTHVMHVTISYPLPD-NQLLTQLFTAL-EEESVNVDMISQIVNL 322
Qy 300 RGRQ-----DRHHLHLLPQTSPPPPWKNWTRSETRSASTOLLYDDDHGKVSLLIGAGM 351
Db 323 EGLQSFSEIKSDAHQISSILE-----NLSTHFSALDYKINEAYVKISLIGSGM 371
Qy 352 RSHPGVTATFCEALAAVGVNIELISTSE 379
Db 372 RDMGVASKAFTTLINSDIPFYQTTTSE 399
```

Search completed: November 21, 2003, 16:11:37
Job time : 9.9415 secs

This Page Blank (uspto)

16	306.5	14.0	439	10	US-09-890-813-16	Sequence 16, Appl
17	285.5	13.0	172	15	US-10-226-136-7	Sequence 7, Appl
18	241	11.0	440	15	US-10-022-832-34	Sequence 34, Appl
19	231	10.5	281	10	US-09-890-813-12	Sequence 12, Appl
20	179	8.2	304	15	US-10-156-761-14293	Sequence 14293, A
21	140	6.4	243	10	US-09-738-626-5718	Sequence 5718, Ap
22	138.5	6.3	294	10	US-09-738-626-6962	Sequence 6962, Ap
23	135	6.2	240	11	US-09-966-521-34	Sequence 34, Appl
24	135	6.2	240	12	US-10-429-094-34	Sequence 40, Appl
25	135	6.2	240	12	US-10-138-701-40	Sequence 88, Appl
26	135	6.2	248	11	US-09-966-521-88	Sequence 88, Appl
27	135	6.2	248	12	US-10-429-094-88	Sequence 60, Appl
28	135	6.2	268	10	US-09-925-537-60	Sequence 60, Appl
29	135	6.2	268	15	US-10-084-205-60	Sequence 239, Appl
30	126	5.7	376	10	US-09-712-363-239	Sequence 239, Appl
31	125	5.7	252	15	US-10-156-761-10162	Sequence 10162, A
32	122.5	5.6	114	10	US-09-890-813-10	Sequence 10, Appl
33	119	5.4	523	11	US-09-880-505-114	Sequence 114, App
34	119	5.4	523	14	US-10-051-643-114	Sequence 114, App
35	119	5.4	541	11	US-09-880-505-160	Sequence 160, App
36	119	5.4	541	12	US-10-205-979-40	Sequence 40, Appl
37	119	5.4	541	14	US-10-051-643-160	Sequence 160, App
38	114	5.2	3798	14	US-10-014-717-6	Sequence 3, Appl
39	113	5.2	540	15	US-10-046-649-3	Sequence 276, App
40	111.5	5.1	163	10	US-09-881-752A-276	Sequence 6, Appl
41	108.5	4.9	540	10	US-09-847-637B-6	Sequence 6, Appl
42	108.5	4.9	540	10	US-09-712-363-169	Sequence 169, App
43	108.5	4.9	540	15	US-10-267-311-4	Sequence 4, Appl
44	108.5	4.9	540	15	US-10-046-649-4	Sequence 4, Appl
45	108.5	4.9	639	15	US-10-267-311-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-793-306-163
; Sequence 163, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 163
; LENGTH: 421
; TYPE: PRI
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: MO-2 (aspartokinase)
US-09-793-306-163

Query Match 73.7%; Score 1615.5; DB 9; Length 421;
Best Local Similarity 85.1%; Pred. No. 1.9e-146; Gaps 3;
Matches 338; Conservative 7; Mismatches 35; Indels 17;
Qy 1 VQYGGSSVADAEIRRAEIRVATKQGNQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
Db 5 VQYGGSSVADAEIRRAEIRVATKQGNQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 64
Qy 61 ELDMLTAGERISNALVMAIESLGAHARSFTCSQAGVITGTHGNAKIIDVTFGLQTA 120

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 16.537 Seconds
(without alignments)
4680.740 Million cell updates/sec

US-09-688-672A-4

Title: Perfect score: 2193
Sequence: 1 VQYGGSSVADAEIRRAEIRVATKQGNQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666198 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1615.5	73.7	421	9	US-09-793-306-163
2	1224.5	55.8	421	10	US-09-738-626-3781
3	1224.5	55.8	421	11	US-09-746-660A-24
4	1223.5	55.8	421	15	US-10-226-136-5
5	1217.5	55.5	421	15	US-10-067-974-2
6	1211.5	55.2	421	15	US-10-067-974-16
7	1144.5	52.2	430	15	US-10-156-761-12093
8	770	35.1	156	12	US-10-084-843-227
9	770	35.1	156	12	US-10-193-002-222
10	673	30.7	394	12	US-09-882-227-290
11	377	17.2	564	10	US-09-890-813-17
12	356	16.2	555	10	US-09-890-813-8
13	350	16.0	560	10	US-09-890-813-6
14	314.5	14.3	262	12	US-10-301-997-23
15	310.5	14.2	449	11	US-09-989-339-23

Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGELQTA 124
 QY 121 LEEGRVVLVAGFGVGSQDTHKDVTTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGSADP 180
 Db 125 LEEGRVVLVAGFGVGSQDTHKDVTTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGSADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRPGTVVVG 240
 Db 185 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRPGTVVVG 244
 QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDI PGYAAKVFRAVARRRRHGAEE----- 295
 Db 245 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDI PGYAAKVFRAVA-----DADVNIDM 295
 QY 296 RLQGRGRQRHLLHPQS---GPPPWKNWTRSETRASTOLLYDDHIGKVSLLIGAGMR 352
 Db 296 VLQNVSKVEDGKTIDTFTCSRDPGPAAVKLSLNRNEIGFSQLLYDDHIGKVSLLIGAGMR 355
 QY 353 SHPGVTATFCEALAAVGVNIELISTSEDQRSCCAAT 389
 Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 2

US-09-738-626-3781

; Sequence 3781, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENO, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OKAZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; PRIOR FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 3781

; LENGTH: 421

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-3781

Query Match 55.8%; Score 1224.5; DB 10; Length 421;

Best Local Similarity 64.0%; Pred. No. 6.9e-109;

Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKGNDVVVVVSAMGTTTDLDDLAQOVCPAPP 60

Db 5 VKYGGSSLESAERIRNVAERIVATKAGNDVVVVCSAMGTTTDELLEAAVNPVPP 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGELQTA 120

Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNARIVDVTPGVR 124

QY 121 LEEGRVVLVAGFGVGSQDTHKDVTTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGSADP 180

Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDITAVAMAAALNADVCEIYSDVDGVYTADP 184

QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRPGTVVVG 240
 Db 185 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYDRPGTVVVG 244
 QY 241 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDI PGYAAKVFRAVA-----RRRRQHR 290
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLSDKPGSEAKVFPALADAINIDMVLQNVSSVE 304
 QY 291 HGAERLQGRQD-RHLLHLP--QTSGPPPWKNWTRSETRASTOLLYDDHIGKVSLLI 347
 Db 305 DGTDTIDTFTCSRDPGPAAVKLSLNRNEIGFSQLLYDDHIGKVSLLI 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 Db 351 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 382

RESULT 3

US-09-746-660A-24

; Sequence 24, Application US/09746660A

; Publication No. US20030049804A1

; GENERAL INFORMATION:

; APPLICANT: Pompejus, Markus

; APPLICANT: Kroger, Burkhard

; APPLICANT: Schroder, Hartwig

; APPLICANT: Zelder, Oskar

; APPLICANT: Haberkauer, Gregor

; APPLICANT: Kim, Jun-Won

; APPLICANT: Lee, Heung-Schick

; APPLICANT: Hwang, Byung-Joon

; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING

; FILE REFERENCE: BGI-121CP2

; CURRENT APPLICATION NUMBER: US/09/746,660A

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 09/606740

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: 09/603124

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: 60/141031

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: 60/142101

; PRIOR FILING DATE: 1999-07-02

; PRIOR APPLICATION NUMBER: 60/148613

; PRIOR FILING DATE: 1999-08-12

; PRIOR APPLICATION NUMBER: 60/187970

; PRIOR FILING DATE: 2000-03-09

; PRIOR APPLICATION NUMBER: DE 19931420.9

; PRIOR FILING DATE: 1999-07-08

; NUMBER OF SEQ ID NOS: 125

; SOFTWARE: PatentIn vers. 2.0

; SEQ ID NO 24

; LENGTH: 421

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-09-746-660A-24

Query Match 55.8%; Score 1224.5; DB 11; Length 421;

Best Local Similarity 64.0%; Pred. No. 6.9e-109;

Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKGNDVVVVVSAMGTTTDLDDLAQOVCPAPP 60

Db 5 VKYGGSSLESAERIRNVAERIVATKAGNDVVVVCSAMGTTTDELLEAAVNPVPP 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGELQTA 120

Db 65 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNARIVDVTPGVR 124

QY 121 LEEGRVVLVAGFGVGSQDTHKDVTTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGSADP 180

Db 125 LDEGKICIVAGFGVGNKTRDVTTLGRGSGDITAVAMAAALNADVCEIYSDVDGVYTADP 184

QY 291 HGAERLQGRQD-RHHLLHP--QTSPPPWKNWTRSETRSASTQLLYDDHIGKVSLL 347
 DB 305 DGTDTITFTCPRADGRAMEILKKLVQVQ-----NW-----TNVLYDDQVKKVSLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISISE 382

RESULT 6

US-10-067-974-16
 ; Sequence 16, Application US/10067974
 ; Publication No. US20030055232A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Li, Ling-Yew
 ; APPLICANT: Irei, Kelli J.
 ; TITLE OF INVENTION: Polynucleotide Constructs for Increased Lysine Production
 ; FILE REFERENCE: 1533.2640001
 ; CURRENT APPLICATION NUMBER: US/10/067,974
 ; CURRENT FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: 60/267,183
 ; PRIOR FILING DATE: 2001-02-08
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 16
 ; LENGTH: 421
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Consensus Sequence of Protein Sequence Alignment
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (40)..(40)
 ; OTHER INFORMATION: May be either Cys or Val
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (317)..(317)
 ; OTHER INFORMATION: May be either Ser or Ala
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (345)..(345)
 ; OTHER INFORMATION: May be either Gly or Asp
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (380)..(380)
 ; OTHER INFORMATION: May be either Thr or Ile
 ; US-10-067-974-16

Query Match 55.2%; Score 1211.5; DB 15; Length 421;
 Best Local Similarity 63.5%; Pred. No. 1.2e-107; Mismatches 61; Indels 27; Gaps 5;
 Matches 249; Conservative 55;
 QY 1 VQYGGSSVADAERIRVAERIVATKKQGNVVVVVSAMGDTTDDLLDLAQVCPAPPP 60
 DB 5 VQYGGSSLSAERIRVAERIVATKKAGNDVVVXSAMGDTTDELLELAANVPVPPAR 64
 QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITGTGHNAKIIDVTPGRLOTA 120
 DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITGTGHNARIVDVTGVRREA 124
 QY 121 LEEGRVVLVAGFGQVSQDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFPSADP 180
 DB 125 LDEGKICIVAGFGQVKNKEDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYVADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNI PVHVRSSYSRDRPGTVVGS 240
 DB 185 RIVPNAOKLEKLSFEEMLEMAAGSKVLRLSRVEYARFNPVRVRSYSDNPGTLLIAGS 244
 QY 241 IKDVPMDPILTVGAHRSSEAKVTIVGLPDPGVAARVAVRAV-----RRRQRH 290
 DB 245 MEDIPVEAVLTGVATDKSAKVTVLGIDSKPGEAAKVFEALADAEINIDVILQNVSSVE 304
 QY 291 HGAERLQGRQD-RHHLLHP--QTSPPPWKNWTRSETRSASTQLLYDDHIGKVSLL 347

DB 305 DGTDTITFTCPRADGRAMEILKKLVQVQ-----NW-----TNVLYDDQVKKVSLV 350
 QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISISE 382

RESULT 7

US-10-156-761-12093
 ; Sequence 12093, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 12093
 ; LENGTH: 430
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-12093

Query Match 52.2%; Score 1144.5; DB 15; Length 430;
 Best Local Similarity 60.2%; Pred. No. 3.4e-101; Mismatches 75; Indels 29; Gaps 4;
 Matches 239; Conservative 54;
 QY 1 VQYGGSSVADAERIRVAERIVATKKQGNVVVVVSAMGDTTDDLLDLAQVCPAPPP 60
 DB 5 VQYGGSSVADAERIKRVAERIVATKKQGNVVVVVSAMGDTTDELIDLAEOVSPMESGR 64
 QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITGTGHNAKIIDVTPGRLOTA 120
 DB 65 EFDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITGDSVHNKARIIDVTPGRIRTA 124
 QY 121 LEEGRVVLVAGFGQVSQDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFPSADP 180
 DB 125 LDEGNIAIVAGFGQVSQDKDITTLGRGSDTTAVALAALDAEVCIEIYTDVGVFTADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNI PVHVRSSYSRDRPGTVVGS 240
 DB 185 RVKKAKKIDWIAFEDMLELAAGSKVLLHRCVEYARRINIPHVRSSFGSGLOGTWV--- 241
 QY 241 IKDVPMDPILTVGAHRSSEAKVTIVGLPDPGVAARVAVRAVRRR--- 286
 DB 242 -SNTPLVQKAOQCEQVEQAIISGVAHDTSEAKVTIVGVDPKPGEAASIFRAIDAENV 300
 QY 287 ----RQRHGAERLQGRQDRHLLHLLPQTSGPPPWKNWTRSETRSASTQLLYDDHIG 342
 DB 301 IDMVQNVNSAASGTGLTDIS-----FTLPKTEGRKAIDALEKAKSVIGFDSLRVDDQIG 353
 QY 343 KVSLLIGAMRSHPGVTATFCEALAAVGVNIELISTSE 379
 DB 354 KISLVGAGMKNTPGVTAGFPEALSDAGVNIELISTSE 390

RESULT 8

US-10-084-843-227
 ; Sequence 227, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843

FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 227:

SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids

TYPE: amino acid
STRANDEDNESS: single

MOLECULE TYPE: protein
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 227:

US-10-084-843-227

Query Match 35.1%; Score 770; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.4e-66;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QKYGSSVADAEIRRVAEIRVATKKQNDVVVWSAMGDTTDDLLDLAQVCPAPPRE 61
Db 1 QKYGSSVADAEIRRVAEIRVATKKQNDVVVWSAMGDTTDDLLDLAQVCPAPPRE 60
QY 62 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 121
Db 61 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 120
QY 122 EGRVVLVAGFCQVSQDTKDVTTLGRGSGDITAVAM 157
Db 121 EGRVVLVAGFCQVSQDTKDVTTLGRGSGDITAVAM 156

RESULT 9
US-10-193-002-222

Sequence 222, Application US/10193002
Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.

Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 222:

SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids

TYPE: amino acid
STRANDEDNESS: single

MOLECULE TYPE: protein
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 222:

Query Match 35.1%; Score 770; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.4e-66;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QKYGSSVADAEIRRVAEIRVATKKQNDVVVWSAMGDTTDDLLDLAQVCPAPPRE 61
Db 1 QKYGSSVADAEIRRVAEIRVATKKQNDVVVWSAMGDTTDDLLDLAQVCPAPPRE 60
QY 62 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 121
Db 61 LDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLOTAL 120
QY 122 EGRVVLVAGFCQVSQDTKDVTTLGRGSGDITAVAM 157
Db 121 EGRVVLVAGFCQVSQDTKDVTTLGRGSGDITAVAM 156

RESULT 10

US-09-882-227-290

Sequence 290, Application US/09882227
Publication No. US20030158396A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20030158396A Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 290
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-290

Query Match 30.7%; Score 673; DB 12; Length 394;
Best Local Similarity 40.1%; Pred. No. 5.8e-56;
Matches 151; Conservative 83; Mismatches 119; Indels 24; Gaps 6;

Qy 13 ERIRVAERIVATKQGNDDVVVVVSAMGDTTDDLDAQQVCPAPPRELDMLLTAGERI 72
Db 5 ERHNVQRVLESVTIQLQVHVVSMSGETDRLEEFQKFNHNPKNKREMDRIVSGELV 64
Qy 73 SNALVAMAIESLGAHARSFTGSAQVITTTGHNKAKIIDVTPCRLQALBEGRVVLVAGF 132
Db 65 SSAALSVALERYGHRALSLSGKEAGILTSHFQNAVIGSIDTKRITELLEKNYIVVIAGF 124
Qy 133 QGVSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFISADPRIVNARKLDTV 192
Db 125 QG-ADTQGETTTLGRGSDLSVALAGALKHCEIYTDVGVITTDPRIBKAKIAQI 183
Qy 193 TFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDRPGTVVVG--SIKDVPMEDPI 250
Db 184 SYDMLSLAMGAKVLLNSVELAKXSVKLVTRNSFNHSEGLTVAEKDFKGERMETPI 243
Qy 251 LTGVANDRSEAKTIVGLDPIGYAAKVFRAVARRRQRHGAERLQGRQRDHLHL 310
Db 244 VSGIALDKQARVSMGEVDREPGIAAEIFGALA---EYRINVDMIVQITGDKTDLDF 299
Qy 311 -LPQSGPPPKWTRSETRASTQL-----YDDHIGKVSLLIGAGMRSHPGVTATFC 362
Db 300 TIVKTO-----TEETKQALKPFLAQWSDIDYENAKVISVGVGMKSHSGVASTAF 350
Qy 363 ELAAAVGVNIELSTSE 379
Db 351 KALAKDNINIMISTSE 367

RESULT 11
US-09-890-813-17
; Sequence 17, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/172944
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Glycine max
US-09-890-813-17

Query Match 17.2%; Score 377; DB 10; Length 564;
Best Local Similarity 27.0%; Pred. No. 2.6e-27;
Matches 128; Conservative 71; Mismatches 165; Indels 110; Gaps 13;

Qy 1 VQYGGSSVADAEIRRAERIVATKKQGNDDVVVVVSAMGDTTDDL-----47

Db 87 VMKFGSSVASADRMKEVATLILSFPEE--RPVIVLSAMGKTTNKLLLAGKAVSCGVIN 144
Qy 48 -----DLAQQVCPAPPREL-----DMLLTAG 69
Db 145 VSSTEEELCFIKDLHLRTVDQLGVDGVSISKHLEQLLKGIAAMKELTKRTQDYLVSFG 204
Qy 70 ERISNALVAMAIESLGAHARSFTGSAQVITTTGHNKAKIIDVT---FCRLQ-TALEEG 124
Db 205 ECMSTRIFAAYLNKIGVKARQYDAFEIGFITTTDDFNADILEATYPAVAKRLHGDWLSDP 264
Qy 125 RVVLVAGFQGVSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFISADPRIVR 184
Db 265 AIAIVTGLGKARKSCAVTTILGRGSDLTATTIGKALGLPEIQVWKDVGDLTCDNIIYP 324
Qy 185 NARKLDTVTFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDR-PGTVVVGSIKD 243
Db 325 KAEPVPLYTDEAAELAYFGAQLHPSMRPARESDIPVRKNSYNPKAPGTLIT---KA 381
Qy 244 VPVEDPILTGVAHDRSEAKTIVGLDPIP--GYAAKVFRAVARRRRQHRHGAERLQGRG 301
Db 382 RDMSKAVLTISVLKRNVTMLDIASTRMLGOYGLAKVFSIFEELGISVDVATSEVSVSL 441
Qy 302 QDRHHLHLHPQSGPPPKWTRSETRASTOLLVDDHIGK-----VS 345
Db 442 TLD-----PSKLWSRELITQASL---DHVVEELEKIAVNVLLQNSIIS 483
Qy 346 LIGAGMRSHPGVTATFCEALAAVGVNIELIS-----TSEBQRGCCAA 388
Db 484 LIGNVQRSSL-ILERLSRVLTGLVTQVQMISQASKVNSLVNDSEAEQVRA 536

RESULT 12
US-09-890-813-8
; Sequence 8, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/172944
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Zea mays
US-09-890-813-8

Query Match 16.2%; Score 356; DB 10; Length 555;
Best Local Similarity 29.9%; Pred. No. 2.7e-25;
Matches 101; Conservative 57; Mismatches 116; Indels 64; Gaps 7;

Qy 1 VQYGGSSVADAEIRRAERIVATKKQGNDDVVVVVSAMGDTTDD-----44
Db 86 VMKFGSSVSAARMAEVAGLITTFPEE--RPVVVLSAMGKTTNNLLLAGKAVGCGVH 143
Qy 45 -----DLLDLAQQVCPA-----PPPRELDMLLTAG 69
Db 144 VSETEENWNVKSLHIKTVDELGLPRSVIQDMLDELEQLLKGIAAMKELTPRTSDYLVSFG 203
Qy 70 ERISNALVAMAIESLGAHARSFTGSAQVITTTGHNKAKIIDVTPCRLQAL-----EEG 124
Db 204 ECMSTRIFAAYLNKIRVKARQYDAFDIGFITTTDEFGNADILEATYPAVAKRLHGDWITQDP 263
Qy 125 RVVLVAGFQGVSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFISADPRIVR 184
Db 264 AIPVVTGLGKWKSGAVTTILGRGSDLTATTIGKALGLREIQVWKDVGDLTCDNIIYP 323
Qy 185 NARKLDTVTFEEMLEMAACGAKVLMRCVYARHNPVHVRSSYSDR-PGTVVVGSIKD 243

```

; Publication No. US20030148346A1
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/301,997
; FILING DATE: 22-No. US20030148346A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,431
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/887,534
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-301-997-23

Query Match 14.3%; Score 314.5; DB 12; Length 262;
Best Local Similarity 33.9%; Pred. No. 8.7e-22;
Matches 87; Conservative 47; Mismatches 82; Indels 41; Gaps 8;

Qy 140 KDVTTILGRGSDPTAVAMAALGADVCEIYTDVDFGSADPRIVRNARKLDTVTREEMLE 199
Db 5 QELTTLGRGSDTTAVALAVS-NQIPCEIYTDVGVATDPRLLPKAKRELDIVSYEEMW 63
Qy 200 MAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGSIKDVPMEDPILTVGAHRS 259
Db 64 MSALGAGVLETSVELAKVNIPLYLGLKTLNVKGTWIMSN--EELEKKAIVTGVLDKH 121
Qy 260 EAKVTI-VGLPDIPGVAKVFAVARRRQHRHGAEE-----FLQR-----GROD 304
Db 122 MMHVTISYFLPD-----NQLLTQLFTEBEGAVNVDMSIQVNLDDLQSLFTTKDSR 173
Qy 305 RHHLHLLPQTSGPPPKWNTSRSTRASATQLLY--DDHTGKVSILGAGMRSHPGVTATFC 362
Db 174 FHOISMLT-----LNQYEAALAYKINEHYVKISLIGSGNRDMSGVASKAF 220
Qy 363 EALAAVGVNIELISTSE 379
Db 221 LTLIENNIPFYQTTTSE 237

RESULT 15
US-09-989-339-23
; Sequence 23, Application US/09899339
; Publication No. US2003008886A1
; GENERAL INFORMATION:

```

```

; Publication No. US20030148346A1
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/301,997
; FILING DATE: 22-No. US20030148346A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,431
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/887,534
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-301-997-23

Query Match 16.0%; Score 350; DB 10; Length 560;
Best Local Similarity 26.4%; Pred. No. 1e-24;
Matches 120; Conservative 76; Mismatches 155; Indels 104; Gaps 12;

Qy 1 VOKYGGSSVADEAIRRVAERIVATKQGNVVVVVSAMGDTTDDLLDLAQ----- 52
Db 86 VMKFGSSVSAARMAEAVGLILTPPE--RPVVVLSAMGKTTNNLLIAGEKAVGCGVIH 143
Qy 53 -----VCPAP-----PPRELDMLLTAGE 70
Db 144 VSEIEEWMNVKSLHKITVDELGLXICNTSYELEQLLKGIAWKELTPRTSDVLSFGE 203
Qy 71 RISNALVAMATIESGAHARSTGSGAVITTTGTHGNAKIIDVTPCRLQAL-----BEGR 125
Db 204 CMSTRIFSAYLKIRVKARQYDAPDIGFTTDFEGNADILEATYPAVAKRLHGDWICDPA 263
Qy 126 VVLVAGFGVQSDTKDVTTLGRGSDPTAVAMAALGADVCEIYTDVDFGSADPRIVRN 185
Db 264 IPVVTGFLGKWKSGAVTTLGRGSDLTATTIGKALGLREIQWKVDVGLTCDPNIYPH 323
Qy 186 ARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDR-PGTVVVGSIKDV 244
Db 324 AKTVPYLTFEEATELAYFGAQLHPQSMRPAEGDIPRVKNSYNPKAPGTLITRQ-RDM 382
Qy 245 PMEDPILTVGAHRSSEAKVTIVGLPDIPGVAKVFAVARRRQHRHGAERLOGRQD 304
Db 383 DXGLVLTLSIV---LKSNTVMD-----IVSTRMLGQYGFARVSGICVIE 425
Qy 305 RHHLHL-----LPQTSGPPPKWNTSRSTRASATQLLYDDHLGK----- 343
Db 426 DLCTISVDCVATSEVSVSLDPSKINSRELIIQASL-----DHVVELEKIAIVRLIQOR 481
Qy 344 --VSLIGAGMRSHPGVTATFCEALAAVGVNIELIS 376
Db 482 AIISLIGNVEQSSLIEXT-GRVLRKSGVNVQMIS 515

RESULT 14
US-10-301-997-23
; Sequence 23, Application US/10301997

```

; APPLICANT: Falco, Saverio Carl
; APPLICANT: Famodu, Layo
; APPLICANT: Rafalski, Jan A.
; APPLICANT: Ramaker, Michael
; APPLICANT: Tarczynski, Mitchell C.
; APPLICANT: Thorpe, Catherine
; TITLE OF INVENTION: PLANT METHIONINE SYNTHASE GENE AND METHODS FOR INCREASING THE
; FILE OF INVENTION: METHIONINE CONTENT OF THE SEEDS OF PLANTS
; FILE REFERENCE: BB-1067-B
; CURRENT APPLICATION NUMBER: US/09/989,339
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 08/703,829
; PRIOR FILING DATE: 1996-08-27
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-989-339-23

Query Match 14.2%; Score 310.5; DB 11; Length 449;
Best Local Similarity 30.2%; Pred. No. 4.6e-21;
Matches 102; Conservative 55; Mismatches 118; Indels 63; Gaps 8;
QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVWVVSAMGDTTDDLLDLAQOVCP----- 55
Db 6 VSKFGTGVADFDAMNRSADIVLS---DANVELVLSASAGITNLLVALAEGLEPGERFE 62
QY 56 -----APPPRELDMLLTAGERIS 73
Db 63 KLDAIRNIQFAILERLRYPNVIREIERLENTVLAEAAALATSPALTDLVSHGELMS 122
QY 74 NALVAMAIESLGAHARSFTGSGAGVITTCGTHNAK-----IIDVTPGRIQTALBEGRVVL 128
Db 123 TLLFVEILRERDVQAOQWFDVRKV-METNDRFGRAEFPDIAALAEALQLLPELNELG-LVI 180
QY 129 VAGFQGVSDTKDVTTLGRGSDTTAVAVAAALGADVCEIYTDVDFISADPRIVRNARK 188
Db 181 TQGFIG-SENKGTTLGRGSDYTHALLAEALHARSVDIWTDFGIYTTDPRVVSAAKR 239
QY 189 LDTVTEEMLEMAACGAKVLMRCVEYARRHNPVHRSSYSYSDRPGTVVVGSIKDVPMED 248
Db 240 IDEIAFAEAAEAATFCAGVLHPATILPAVRSDIPVFGSSKDPKPRAGGTLVCNKTEPN--- 296
QY 249 PILTGVADRSEAKVTIVGLPDI--PGYAAKVRAVAR 284
Db 297 EUFRALALRNQTLTLHSLNMLHSGFLAEVFGILAR 334

Search completed: November 21, 2003, 16:38:11
Job time : 18.537 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 8.84535 Seconds
(without alignments)

4609.825 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VQYGGSSVADAEIRVAE.....SAATRRPCTRGDRGWACQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1615.5	73.7	421	2 F70794	probable ask prote
2	1545.5	70.5	421	2 G87199	aspartokinase I mp
3	1478.5	67.4	421	2 S82432	aspartate kinase (
4	1229.5	56.1	421	2 S15276	aspartate kinase (
5	1141	52.0	425	2 T35383	probable aspartoki
6	823	37.5	412	2 C83531	aspartate kinase a
7	816.5	37.2	600	2 S76764	hypothetical prote
8	785	35.8	411	2 A48946	aspartate kinase (
9	778	35.5	407	2 J24640	aspartate kinase (
10	776.5	35.4	606	2 A22611	aspartate kinase (
11	761	34.7	405	2 H81865	aspartate kinase (
12	760	34.7	405	2 F81076	aspartate kinase (
13	757.5	34.5	412	2 H84036	aspartokinase, alp
14	742.5	33.9	400	2 E81405	aspartokinase II a
15	734.5	33.5	415	2 B70399	aspartokinase - Aq
16	724.5	33.0	401	2 G72245	aspartokinase II -
17	724.5	33.0	739	2 H72364	aspartokinase II -
18	721	32.9	405	2 H71843	aspartokinase 2 al
19	712.5	32.5	408	2 A29314	aspartate kinase (
20	711	32.4	405	2 E64673	aspartokinase - He
21	707	32.2	404	2 A81582	aspartokinase II a
22	705	32.1	415	2 H87353	aspartokinase I mp
23	703	32.1	404	2 A81229	aspartokinase II a
24	694.5	31.7	423	2 A83275	aspartate kinase (
25	656	29.9	431	2 AF3069	aspartate kinase (
26	656	29.9	431	2 D38217	aspartate kinase (
27	632	28.8	395	2 A82311	aspartokinase, alp
28	585	26.7	401	2 A89916	aspartokinase II l
29	562	25.6	411	2 H83949	aspartokinase I (a

RESULT 1

F70794

Probable ask protein - Mycobacterium tuberculosis (strain H37Rv)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: F70794

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70794

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <COL>

A;Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA18031.1; PID:g2960133

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: ask

C;Superfamily: aspartate kinase; aspartate kinase homology

F;3-406/Domain: aspartate kinase homology <DKI>

Query Match 73.7%; Score 1615.5; DB 2; Length 421;
Best Local Similarity 85.1%; Pred. No. 2.2e-112;
Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;

Qy 1 VQYGGSSVADAEIRRVAAERIVATKKQNDVVVVVVSAMGDTTDDLDDLAQQVCPAPPR 60

Db 5 VQYGGSSVADAEIRRVAAERIVATKKQNDVVVVVVSAMGDTTDDLDDLAQQVCPAPPR 64

Qy 61 ELDMLLTAGERISNALVMAIESLGHARSFTGSGQNGVITTTGTHGNAKIIDVTPGLQTA 120

Db 65 ELDMLLTAGERISNALVMAIESLGHARSFTGSGQNGVITTTGTHGNAKIIDVTPGLQTA 124

Qy 121 LEBGRVVLVAGFGVGSQDTKDTTLGRGSGDITAVAMAAALGADVCEIYTDVDGIFSA DP 180

Db 125 LEBGRVVLVAGFGVGSQDTKDTTLGRGSGDITAVAMAAALGADVCEIYTDVDGIFSA DP 184

Qy 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRRNI PVHVRSSVSDRPGTVVGS 240

Db 185 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRRNI PVHVRSSVSDRPGTVVGS 244

Qy 241 IKDVPNEDPILGVANDRSEAKVTIVGLPDIPCYAAKVFARARRRQRHGAEE----- 295

Db 245 IKDVPNEDPILGVANDRSEAKVTIVGLPDIPGYAAKVFARA-----DADVNIDM 295

Qy 296 RLQGRGRQRHLLHPOTS---GPPPWKNWTRSETRSASTQLLYDDHIGKVSILGAGMR 352

Db 296 VLQNVKVEDGKTDITFTCRDVGPAAVEKLSLRNEIGFSQLLYDDHIGKVSILGAGMR 355

ALIGNMENTS

```

30      532      24.3      404      2      C46665      aspartate kinase (
31      524      23.9      403      2      AB1617      aspartokinase I (a
32      520      23.7      401      2      D97845      aspartate kinase (
33      506      23.1      403      2      AD1254      aspartokinase I (a
34      492      22.4      473      2      C64371      aspartate kinase (
35      480.5      21.9      446      2      E71635      aspartokinase I (ys
36      462      21.1      399      2      D97123      aspartokinase I mp
37      440.5      20.1      462      2      D69337      aspartate kinase (
38      440      20.1      473      2      F75405      aspartate kinase -
39      391.5      17.9      819      2      AH0056      aspartate kinase (
40      377      17.2      569      2      T48575      aspartate kinase -
41      370.5      16.9      820      2      AC0502      aspartokinase I/ho
42      369.5      16.8      820      1      DESCK      chrA Bifunctional
43      369.5      16.8      820      2      B90629      aspartokinase I-ho
44      369.5      16.8      820      2      B85480      aspartokinase I, h
45      366.5      16.7      825      2      H82086      aspartokinase I/ho

```

QY 353 SHPGVTATFCEALAAVGVNIELISTSEDORSRCCAAAT 389
 |||||
 Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392
 |||||

RESULT 2

G87199
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 14-Dec-2001
 C:Accession: G87199
 R: Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,
 R.; Davies, R.M.; Devlin, K.; Duchov, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; MUID:21128732; PMID:11234002
 A:Accession: G87199
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-421 <STO>
 A:Cross-references: GB:AL450380; NID:G13093941; PIDN:CAC31839.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ask
 C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 70.5%; Score 1545.5; DB 2; Length 421;
 Best Local Similarity 80.4%; Pred. No. 3.5e-107;
 Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 60
 |||||
 Db 5 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 64
 |||||

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 120
 |||||
 Db 65 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 124
 |||||

QY 121 LEEGRVVLVAGFGQVSQDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGSADP 180
 |||||
 Db 125 LEEGRVVLVAGFGQVSQDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGSADP 184
 |||||

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSDRPGTVVGS 240
 |||||
 Db 185 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSDRPGTVVGS 244
 |||||

QY 241 IKDVPNEDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 295
 |||||
 Db 245 IKDVPNEDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----DADVNIDM 295
 |||||

QY 296 RLQGRQRDRHHLHL---LPQSGPPPKWTRSETRSASTOLLYDDHIGKVSIGAGMR 352
 |||||
 Db 296 VLQNSKVEDGKTDITFTCSRDSGPIAVAKLSLRDEIGFTQLLYDDHIGKVSIGAGMR 355
 |||||

QY 353 SHPGVTATFCEALAAVGVNIELISTSEDORSRCCAAAT 389
 |||||
 Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392
 |||||

RESULT 3

S42422
 C:Species: Mycobacterium smegmatis
 C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 18-Jun-1999
 C:Accession: S42422; S31801
 R: Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
 Mol. Microbiol. 11, 629-639, 1994
 A:Title: Isolation and characterization of the aspartokinase and aspartate semialdehyde
 A:Reference number: S42421; MUID:194254720; PMID:7910936
 A:Accession: S42422
 A:Molecule type: DNA

A:Residues: 1-421 <CIR>
 A:Cross-references: EMBL:Z17372; NID:G44506; PIDN:CAA78994.1; PID:G581352
 A:Note: the authors translated the initiation codon GTG for residue 1 as Val
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: aspartate kinase; aspartate kinase homology
 C:Keywords: alternative initiators; phosphotransferase
 F:1-421/Product: aspartate kinase alpha chain #status predicted <ASA>
 F:3-406/Domain: aspartate kinase homology <DKI>
 F:250-421/Product: aspartate kinase beta chain #status predicted <ASB>

Query Match 67.4%; Score 1478.5; DB 2; Length 421;
 Best Local Similarity 77.6%; Pred. No. 3.3e-102;
 Matches 302; Conservative 31; Mismatches 35; Indels 21; Gaps 4;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 60
 |||||
 Db 5 VKYGGSSVADAEIRIRVAERIVATKKQNDVVVVVSAMGDTTDLDDLAQVCPAPP 64
 |||||

QY 61 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 120
 |||||
 Db 65 ELDMLLTAGERISNALVAVAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 124
 |||||

QY 121 LEEGRVVLVAGFGQVSQDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGSADP 180
 |||||
 Db 125 LEEGRVVLVAGFGQVSQDVTTLGRGSGDTTAVAMAALGADVCEIYTDVDFGSADP 184
 |||||

QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSDRPGTVVGS 240
 |||||
 Db 185 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRNIPVHVRSSYSDRPGTVVGS 244
 |||||

QY 241 IKDVPNEDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 295
 |||||
 Db 245 IKDVPNEDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----ADVNIDM 295
 |||||

QY 296 RLQGRQRDRHHLHL---LPQSGPPPKWTRSETRSASTOLLYDDHIGKVSIGAG 350
 |||||
 Db 296 VLQNSKVEDGKTDITFTCSRDSGPIAVAKLSLRDEIGFTQLLYDDHIGKVSIGAG 353
 |||||

QY 351 MRSPGVTATFCEALAAVGVNIELISTSE 379
 |||||
 Db 354 MRSPGVTATFCEALAAVGVNIELISTSE 382
 |||||

RESULT 4

S15276
 C:Species: Corynebacterium glutamicum
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Jun-1999
 C:Accession: S15276; S49977
 R: Kalinowski, J.; Cremer, J.; Bachmann, B.; Eggeling, L.; Sahm, H.; Puehler, A.
 Mol. Microbiol. 5, 1197-1204, 1991
 A:Title: Genetic and biochemical analysis of the aspartokinase from Corynebacterium glut
 A:Reference number: S15276; MUID:92065816; PMID:1956296
 A:Accession: S15276
 A:Molecule type: DNA
 A:Residues: 1-421 <PUE>
 A:Cross-references: EMBL:X57226; NID:G40509; PIDN:CAA40502.1; PID:G580983
 A:Experimental source: ATCC 13032
 R: Serebrijski, I.; Wojcik, F.; Reyes, O.; Leblon, G.
 submitted to the EMBL Data Library, November 1994
 A:Description: Two loci of Corynebacterium glutamicum ATCC17965 that complement Escherich
 A:Reference number: S49977
 A:Accession: S49977
 A:Molecule type: DNA
 A:Residues: 329-421 <SR>
 A:Cross-references: EMBL:X82928; NID:G599716; PIDN:CAA58100.1; PID:G599717
 R: Kalinowski, J.; Bachmann, B.; Thierbach, G.; Puehler, A.
 Mol. Gen. Genet. 224, 317-324, 1990
 A:Title: Aspartokinase genes lycS-alpha and lycC-beta overlap and are adjacent to the as
 A:Reference number: S12250; MUID:91094767; PMID:1980002
 A:Contents: annotation

C;Genetics:
A;Gene: lycC-alpha
A;Start codon: GTG
C;Superfamily: aspartate kinase; aspartate kinase homology
C;Keywords: alternative initiators; phosphotransferase
F;3-406/Domain: aspartate kinase homology <DKI>

Query Match 56.1%; Score 1229.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 1e-83;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 1 VOKYGGSSVADAERIRRAERIVATKKQNDVAVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VOKYGGSSLESAERIRNVAERIVATKKAGNDVAVVVSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIVDTPGRLOTA 120
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIVDTPGRVREA 124

QY 121 LEEGRVVLVAGFGVQSDTKVITLGRGGSDTTAVMAAALGADVCEIYDVGDFISADP 180
DB 125 LDEGKICIVAGFGVQSDTKVITLGRGGSDTTAVMAAALGADVCEIYDVGDFISADP 184

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240
DB 195 RIVPNAQKLEKLSFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 244

QY 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDIPGYAAKVFRARRR-----RRRQHR 290
DB 245 MEDIPVEAVLTGVATDKSEAKVTIVGLPDIPGYAAKVFRARRR-----RRRQHR 290

QY 291 HGAERLQGRQD-RHLLHLLP--QTSGPPPKWNTSRSTRASQTQLLYDDHIGKVSLLI 347
DB 305 DGTDTITTCPRSDGRRAMEILKKLVQV-----NW-----TNVLYDDQVGRKSLV 350

QY 348 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 379
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 5
TJ35383
probable aspartokinase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 31-Jan-2000
C;Accession: TJ35383
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21576
A;Accession: TJ35383
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-425 <MUR>
A;Cross-references: EMBL:AL079348; PIDN: CAB45482.1; GSPDB: GN00070; SC0EDB: SC66T3.26C
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: ask; SC0EDB: SC66T3.26C
C;Superfamily: aspartate kinase; aspartate kinase homology

Query Match 52.0%; Score 1141; DB 2; Length 425;
Best Local Similarity 61.6%; Pred. No. 3.9e-77;
Matches 239; Conservative 54; Mismatches 79; Indels 16; Gaps 4;

QY 1 VOKYGGSSVADAERIRRAERIVATKKQNDVAVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VOKYGGSSVADAERIKRAKRAIVEAKKNGVAVVVSAMGTTDELIDLAEOVSPFAGR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIVDTPGRLOTA 120
DB 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIVDTPGRIRTS 124

QY 121 LEEGRVVLVAGFGVQSDTKVITLGRGGSDTTAVMAAALGADVCEIYDVGDFISADP 180
DB 125 LEEGRVVLVAGFGVQSDTKVITLGRGGSDTTAVMAAALGADVCEIYDVGDFISADP 180

DB 125 VDEGNVAIVAGFGVQSDSKITTLGRGGSDTTAVMAAALDADVCEIYDVGDFVFTADP 184

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240

DB 185 RVVPRAKKIDWISFEDMLELAASGSKVLLHRCVEYARRHNIPIVHVRSSFSGLQGTWVSSE 244

QY 241 -IK--DVPMDPILTCVAHDSREAKVTIVGLPDIPGYAAKVFRARRR-----RQHRH 291

DB 245 PIQGEKHVEQALISGVAHDTSEAKVTIVGVGPKFGEAAAFRAIADAGVNDIMVYQVNS 304

QY 292 GAAERLQGRQD-RHLLHLLP--QTSGPPPKWNTSRSTRASQTQLLYDDHIGKVSLLI 347

DB 305 ALSTGLTDS-----FTLPKSEGRKAIDALEKRPFGIFDLSLYDDQIGKISLVGAGN 357

QY 352 RSHPGVTATFCEALAAVGVNIELISTSE 379

DB 358 KSNPGVTADFTTALS DAGVNIELISTSE 385

RESULT 6
C83531
aspartate kinase alpha and beta chain PA0904 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 14-Dec-2001
C;Accession: C83531
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-412 <STO>
A;Cross-references: GB:AE004525; GB:AE004091; NID:G9946805; PIDN:AAG04293.1; GSPDB:GN001;
A;Experimental source: strain PA01
C;Genetics:
A;Gene: lycC; PA0904
C;Superfamily: aspartate kinase; aspartate kinase homology

Query Match 37.5%; Score 823; DB 2; Length 412;
Best Local Similarity 42.6%; Pred. No. 1.6e-53;
Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;

QY 1 VOKYGGSSVADAERIRRAERIVATKKQNDVAVVVSAMGDTTDDLLDLAQVCPAPPPR 60

DB 5 VOKFGTSGVTVERIEQVAEKVKKFEAGDDVVVVVSAMSGETNRLIGLANQIMEQFVPR 64

QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIVDTPGRLOTA 120

DB 65 ELDVYVSTGEQVTIALLSMALIKRGVPAVSYTGQVRIILTDSATKARILHIDDDTHRAD 124

QY 121 LEEGRVVLVAGFGVQSDTKVITLGRGGSDTTAVMAAALGADVCEIYDVGDFISADP 180

DB 125 LKAGRVVVVAGFGV-DGNGNITTLGRGGSDTTGVALAALKADECIYTDVGVYTTDP 183

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSYSDRPGTVVVG 240

DB 184 RVVPQARRLDKITFEEMLEMAASLGSKVLQIRAVEFACKNVPLRVLHVSFEQEGTILITD 243

QY 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDIPGYAAKVFRARRRQ----- 288

DB 244 DEESMEQPIISGIATFNDEAKLTIRGVPTPGVAFKILGPISAAVNEVDMIVQVNAHDN 303

QY 289 -----HEHGAERLQGRQD-RHLLHLLP--QTSGPPPKWNTSRSTRASQTQLLYDDH 340

DB 304 TTDTFTVVRNDVYNALE-----ILKQTA-----ANIGAKEALGDTN 340

QY 341 IGKVSLLIGAGMRSHPGVTATFCEALAAVGVNIELISTSE 379

DB 341 IAKVISVGVGMRSAGVASRMFEALAKESINIQMISTSE 379

RESULT 7

S76764
 A:Note: sequence modified after extraction from NCBI backbone
 A:Note: sequence extracted from NCBI backbone (NCBIN:117802)
 A:Accession: B48946
 A:Molecule type: protein
 A:Residues: 2-9 <SC2>
 A:Note: this material was purified alpha chain
 A:Accession: C48946
 A:Molecule type: protein
 A:Residues: 246-253 <SC3>
 A:Note: this material was purified beta chain; the beta chain apparently was not blocked
 C:Comment: The amino terminal Met of the beta chain is preceded by a strongly predicted
 codon was demonstrated after deletion of the alpha chain region. This suggests that expres
 s.
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 A:Reference number: S74322; MUID:97061201; PMID:8505231
 A:Accession: S76764
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-600 <KAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8676.1; PID:d101940
 C:Superfamily: aspartate kinase homology
 F:3-417/Domain: aspartate kinase homology <DK1>

Query Match 37.2%; Score 816.5; DB 2; Length 600;
 Best Local Similarity 43.9%; Pred. No. 7.6e-53;
 Matches 183; Conservative 83; Mismatches 118; Indels 33; Gaps 7;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLDAQVCPAPPR 60
 DB 5 VQYFGSTVGTVERIQAVQRIKRTVQGGNSLVVVSAMGKSTDLVLDIAQISNPCCR 64
 QY 61 ELDMLLTAGERISNALVAVMAIESLGHAHSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 DB 65 EMDMLLTGEQVSIALLSLALQIDQPAISLTGAQGVITVEAHSRRARILEIRPDRLEHH 124
 QY 121 LEEGRVLVAGFGVQSQ-DTKQVTLGRGSGDTTAVAMAALGADVCEIYTDVDFISAD 179
 DB 125 LREGKVVVAGFGQISSVEHLLTTLGRGSGDTSVALAALAKADFCETIYDVPGLITTD 184
 QY 180 PRIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVG 239
 DB 185 PRLVPEQLMAEITCDMELESLGAKVLPRAVEIARNYGIPLVRSWSDEPQTKVA 244
 QY 240 SKDVMEDPILT-----GVADRSEAKVTIVGLDPDIFYAAKVFRAVARR----- 286
 DB 245 P-----EVQNRSLVGLSIAKAVDGVEDQAKVALLRVPDRFGVASKLFRDIAQQQVD 300
 QY 287 ---RQHRGGAERLQGRGDRHLLHPQTSGPPPKWTRSETRSASTOLLYDDHIGK 343
 DB 301 LIQSHDGNNDIATVVKDLNLTAEVTSIAIAFALRSYPAQDEA---EIIIVEKGIAK 357
 QY 344 VSLIGAGMESHGPGVTATFCEALAAVGVNIELISTSE-----DQR--SRCAATPN 391
 DB 358 IATAGAGMIGRPGIAAKFKTLADVGVNTEMISTSEVKVSCVIDQORDADRAIALSN 414

RESULT 8

A48946
 A:Note: sequence modified after extraction from NCBI backbone
 A:Note: sequence extracted from NCBI backbone (NCBIN:117802)
 A:Accession: B48946
 A:Molecule type: protein
 A:Residues: 2-9 <SC2>
 A:Note: this material was purified alpha chain
 A:Accession: C48946
 A:Molecule type: protein
 A:Residues: 246-253 <SC3>
 A:Note: this material was purified beta chain; the beta chain apparently was not blocked
 C:Comment: The amino terminal Met of the beta chain is preceded by a strongly predicted
 codon was demonstrated after deletion of the alpha chain region. This suggests that expres
 s.
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 A:Reference number: S74322; MUID:97061201; PMID:8505231
 A:Accession: S76764
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-600 <KAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8676.1; PID:d101940
 C:Superfamily: aspartate kinase homology
 F:3-417/Domain: aspartate kinase homology <DK1>

A:Note: sequence modified after extraction from NCBI backbone
 A:Note: sequence extracted from NCBI backbone (NCBIN:117802)
 A:Accession: B48946
 A:Molecule type: protein
 A:Residues: 2-9 <SC2>
 A:Note: this material was purified alpha chain
 A:Accession: C48946
 A:Molecule type: protein
 A:Residues: 246-253 <SC3>
 A:Note: this material was purified beta chain; the beta chain apparently was not blocked
 C:Comment: The amino terminal Met of the beta chain is preceded by a strongly predicted
 codon was demonstrated after deletion of the alpha chain region. This suggests that expres
 s.
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 A:Reference number: S74322; MUID:97061201; PMID:8505231
 A:Accession: S76764
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-600 <KAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8676.1; PID:d101940
 C:Superfamily: aspartate kinase homology
 F:3-417/Domain: aspartate kinase homology <DK1>

Query Match 35.8%; Score 785; DB 2; Length 411;
 Best Local Similarity 42.1%; Pred. No. 1e-50;
 Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;
 QY 1 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLDAQVCPAPPR 60
 DB 5 VQYFGSTVGTVERIQAVQRIKRTVQGGNSLVVVSAMGKSTDLVLDIAQISNPCCR 64
 QY 61 ELDMLLTAGERISNALVAVMAIESLGHAHSFTGSQAGVITTHGNAKIIDVTPGRLQTA 120
 DB 65 EMDMLLTGEQVSIALLSLALQIDQPAISLTGAQGVITVEAHSRRARILEIRPDRLEHH 124
 QY 121 LEEGRVLVAGFGVQSQ-DTKQVTLGRGSGDTTAVAMAALGADVCEIYTDVDFISAD 180
 DB 125 LREGKVVVAGFGQISSVEHLLTTLGRGSGDTSVALAALAKADFCETIYDVPGLITTD 183
 QY 181 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDPGTVVVG 240
 DB 184 RYVKSARKLASISYDEMLELANLGAGVLPRAVEFAKNGYITLEVSRSSMEREBEGTIE-- 241
 QY 241 IKQVPEM-DPIILGVADHRSSEAKVTIVGLDPDIFYAAKVFRAVARRRQHRHGAERLQ 299
 DB 242 -EVTMEQNJVVGAFEDBITRTVFGLPNSLTSITFTILA----- 284
 QY 300 RGRQDRHLLHPQTSGPPPKWTRSETRSASTOLLYDD----- 339
 DB 285 ---QNRINVDIIQSA-----TDAETNLSFSIKSDDEETMAVENNNKLLNVQGI 333
 QY 340 ---HICKVSLIGAGMESHGPGVTATFCEALAAVGVNIELISTSE 379
 DB 334 ESETGLAKVSIIVSGMISNPGVAAKMFVLAALNGIQVKWVSTSE 377

RESULT 9

JC4640
 A:Note: sequence modified after extraction from NCBI backbone
 A:Note: sequence extracted from NCBI backbone (NCBIN:117802)
 A:Accession: B48946
 A:Molecule type: protein
 A:Residues: 2-9 <SC2>
 A:Note: this material was purified alpha chain
 A:Accession: C48946
 A:Molecule type: protein
 A:Residues: 246-253 <SC3>
 A:Note: this material was purified beta chain; the beta chain apparently was not blocked
 C:Comment: The amino terminal Met of the beta chain is preceded by a strongly predicted
 codon was demonstrated after deletion of the alpha chain region. This suggests that expres
 s.
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 A:Reference number: S74322; MUID:97061201; PMID:8505231
 A:Accession: S76764
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-600 <KAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8676.1; PID:d101940
 C:Superfamily: aspartate kinase homology
 F:3-417/Domain: aspartate kinase homology <DK1>

F;3-397/Domain: aspartate kinase homology <DKI>

Query Match 35.5%; Score 778; DB 2; Length 407;
Best Local Similarity 43.2%; Pred. No. 3.4e-50;
Matches 173; Conservative 77; Mismatches 98; Indels 52; Gaps 7;

Qy		1	VQYGGSSVADAEIRRRVAERIVATKKQGNDVVVVVSAMGTTDDLLDLAQVCAPP	60
			: : : :	
Dd		5	VQKFGGTSGSIERIQHVRNRIEEVQKGNDVVVVVSAMGKTDELNLAKOISNHPSKR	64
			: : : :	

Qy 61 ELDMLLTAGERISNALVMAAIESLGAHARSETGSQAQVITCTTHGNAKIIDVTPEQLQTAL
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
Db 65 EVDMLSTGEOVSTALLANSLHEKGYKAUSLTGWQAQITTEWHGNARIMNIDTRIRRC

Qy 121 LEGRVVLVAGFQGVQSQTVDVTTTCRGGSDDTTAVAAALGADVCEIYTTDVGIFSDP 18

125 IDEGATIVVAGCGQVTFEGAEITTTTCRGGSDDTTAVAAALGAEKCNITVTTGCVFTTDP 18

QY 181 RIYVNRKLDVTTFEEMLEVAACGAKVLMLECEVEYARRHNI PVHRRSYSDRPGTVTVGS 24

Qy 241 IKDVPMDP-ILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRQRRHGAAERLQG 299

Qy 300 RGRQDRHHLLPQTSGPPPWNTSRAS-----TQLL-----YDD 33

DD -----GIVVLLIIQ-----KAINCSEHAEVGLCAIKEDDELELITXZQIIEECLDILMEESS
QY 340 HIGKVSLIGAGMRSHPGVTATFCEALAAVGYNIELISTSE 379
 : |||::|||:|||:|||:|||:|||::|||

DB 334 GLARVSVIGSGMISNPGVAARVFEVLADQGLIENMVASE 373
RESULT 10

AE2261
aspartate kinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AE2261
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.

DNA Res. 8, 205-213, 2001
A; Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Nostoc commune* PCC 7122
A; Reference number: AB1807; MUID:21595285; PMID:11759840

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-606 <KUR>

```

A:Experimental source: strain PCC 7120
 A:Cross-references: GB:BA000003; FJUN:BA035343; FID:G173277; G022:G002
 C:Genetics:
 A:Gene: alr364

Query Match	35.4%;	Score	776.5;	DB 2;	Length	606;			
Best Local Similarity	43.2%;	Pred. No.	7.2e-50;						
Matches	175;	Conservative	74;	Mismatches	115;	Indels	41;	Gaps	

QY 1 VQYGGSSVADAERIRRAERIIVATKKQGNVVVVVSAMGDTTDDLDDLAQQVCAPPPR 60

QY 61 ELDMLITAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPGRLOTA 12

QY 121 LEGRVVLVAGQGVSQD-TKDYTTLGRGSDTTAVAMAALGADVCEIYTDVDGIESAD 17

QY 180 PRIVNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVV- 23

[illegible]

Qy 288 QEHGAERLQRG-----FQDRHHLLPOTSGPPKWNTRSETRSASTQ 334

QY 335 LLYDDHTGKVLIGAGMRSHPGVTATFCEALAAVGVNIELISTSE 379

RESULT 11

H81865
aspartate kinase (EC 2.7.2.4) NVAL701 [imported] - *Neisseria meningitidis* (strain H81865)
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence revision 05-May-2000 #text change 02-Feb-2001

C:Accession: H81865
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; R

A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis*
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81865

A: Molecule type: DNA
A: Residues: 1-405 <PAR>
A: Cross-references: GB:AL162756; GB:AL157959; NID:G7380091; PIDN:CAB84929.1; P

C:Genetics:
A:Gene: *lysc*; NM1701
C:Superfamily: aspartate kinase; aspartate kinase homology

Cf. *keyword*: phosphotransferase

Query Match	34.7%	Score 761	DB 2	Length 405
Best Local Similarity	40.8%	Pred. No. 6.1e-49		

QY 1 VQYGGSSVADAERIRRVARRIVATKKQGNDDVVVVVSAMGTTDDLLDLAQVCPAPPR 60
Matches 161; Conservative 82; Mismatches 116; Indels 30; Gaps

Db	5	VHKYGGTSGSPERIKNAKRVAKARAEGHDIVVWSAMSGETNRLVALAHQEHDPDF	6
QY	61	ELDMLLTAGRISNALVMAIESLGAHARSTGSGAQVITTTGTHGNAKIIDVTPGRIQTA	120

Db	65	ELDVVLSTGEQVTTIGLLAMALKDVGDAKSYTGWQVAKTKDTAHTKARIESIDDERKVRAD	180
QY	121	LEEGSWLVAGFGQVSYQDITKDWTTTLGRGGSDTTAVAMAAALGADVCHIIYTDVVDGIFSDPP	180

Db 125 LTAGKVVIVAGFGQGISSE -GDISTLGRGGSDTSAVALAAALKADECOIYTDVDGVYTTDP 181

QY 181 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRLCVEYARRHNPVHVRSYSYSDRPTVVVGS 240

Db 184 RVWPEARRMDVTTFEMIELASGSKVLQIRSFAGKYKVLRLVSSLQDGGNGTLTF 243

QV 241 IKDVMEDPILTGVAHDRSEAKVTIVGLPDIPGYAAKVFRAVARRRRQHPHGAERLQGR 300

```

Db      244  EEDDNMRAAVTGIAPFDKNQARINVRGVPDKPGVAIQILGAVADANIE-----299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qv      301  GRQDHHIHLIPOTSQPPPPKKNWTRSETRSASTOLLY-----DDHICKV 344

```

```

Db      292  -----VDMIIONVGSEGTDFSFVPRGDYKQTLLEILSERQDSIGAASIDGDDTVCKV 344
               : :: : :: :
Cn      345  SLICGWPSHPCVTAACEALAAAGVNIETSTSE 379

```

Db 345 SAVGLGRSHYGVAAKIFRTLLAEEGINIQMISTSE 379

RESULT 12
F81076

aspartokinase, alpha and beta chains NMB1498 [imported] - *Neisseria meningitidis* (strain C/Species: *Neisseria meningitidis*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: F81076
 R/Tettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, B.R.; Rappuoli, R.; Vitti, A.; Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: F81076
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-405 <TET>
 A/Cross-references: GB:AE002499; GB:AE002098; NID:G7226737; PIDN:AAF41854.1; PID:G722674
 A/Experimental source: serogroup B, strain MC58
 C/Genetics:
 A/Gene: NMB1498
 C/Superfamily: aspartate kinase; aspartate kinase homology

Query Match 34.7%; Score 760; DB 2; Length 405;
 Best Local Similarity 40.8%; Pred. No. 7.3e-49;
 Matches 161; Conservative 81; Mismatches 117; Indels 36; Gaps 3;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 5 VHYGGSVSGSPEIRKNAKRAKARAGHDIVVVVSAMSGETNRLVALAHQEHDPDR 64
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTGTHGNAKIIDVTGRLQTA 120
 Db 65 ELDVLTAGEQVITGLLAWKLDIGVDAKSYTGQVALXTDTAHTKARIESIDDEKMRAD 124
 Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
 Db 125 LTAGKVVIVAGFGIASE-GDSTLGRGGSDTSVALAALKADEGCIYTDVDFGVITDP 183
 Qy 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVWGS 240
 Db 184 RVVPEARRMDTVTFEEMLESLGHSKVLQIRSVFAGKVKVLRVLSSLDQGGNGTLITF 243
 Qy 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDI PGYAAKVFRAVARRRRQHRHGAERLQGR 300
 Db 244 EEDNNVERAAVTGIAFDKQARINVRGPKGVAVQILGAVADANIE----- 291
 Qy 301 GRQDRHLLHLLPQTSPPFPKWNTRSETRSASTQLLY-----DDHIGKV 344
 Db 292 -----VDMIIQNVGSEGTDFSFVPRGDYKQTLLEILSERQDSIGAASIDGDDTVCKV 344
 Qy 345 SLIGAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 345 SAVGLGMRSHVGAAKIFRTLAEBEGINIQMISTSE 379

RESULT 13

H84036

aspartokinase II alpha and beta subunits lysC [imported] - *Bacillus halodurans* (strain C/Species: *Bacillus halodurans*
 C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C/Accession: H84036
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and A/Reference number: A83650; MUID:20512562; PMID:11058132
 A/Accession: H84036
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-412 <STO>
 A/Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA06815.1; GSPDB:GN00
 A/Experimental source: strain C-125
 C/Genetics:
 A/Gene: lysC
 C/Superfamily: aspartate kinase; aspartate kinase homology

Query Match 34.5%; Score 757.5; DB 2; Length 412;
 Best Local Similarity 42.0%; Pred. No. 1.1e-48;
 Matches 173; Conservative 64; Mismatches 102; Indels 73; Gaps 7;
 Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 5 VQYGGTSVSGVQRIEIVANRIIATAASGKXVVVVVSAMGKTTDELVLGLASEITDEPKR 64
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTGTHGNAKIIDVTGRLQTA 120
 Db 65 EMDMLLTAGEQVITALLALHKKGQEQATSLTGQAGNITESHGHNARIIEVERARDEL 124
 Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
 Db 125 LDKVHVIVVAGFGIDQQ-GEIMTLGRGGSDTSVALAALKAEKRCBIYTDVTGVFTDP 183
 Qy 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVWGS 240
 Db 184 RVVQARLQTIISYDEMLEMANIGASVHLHRAVEFAKNIYEIPLLVASSWVDEPGTME-- 241
 Qy 241 IKDVPMDPILTCVAHDSREAKVTIVGLPDI PGYAAKVFRAVARRRRQHRHGAERLQGR 299
 Db 242 -EEGTMEQNLVIRGIAFENDVTKVTHGLP-----QT-SGPPPWKWNTRSETRSASTQLL 336
 Qy 300 RGRQDRHLLHLLP-----YDDHIGKVSLICAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 271 -----NEIHTLPTLFAHLSAAGINVDIIIONQDTSMAISFSITQSSLETARLEELK 324
 Qy 337 -----YDDHIGKVSLICAGMRSHPGVTATFCEALAAVGNIELISTSE 379
 Db 325 GELSFQGDIEEELAKVISVSGMWSNPGVAAKMFELANDIAIKMVTISE 376

RESULT 14

E81405

aspartate kinase (EC 2.7.2.4) alpha and beta chains Cj0582 [imported] - *Campylobacter jejuni* C/Species: *Campylobacter jejuni*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C/Accession: E81405
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrall Nature 403, 665-668, 2000
 A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals type A/Reference number: A81250; MUID:20150912; PMID:10688204
 A/Accession: E81405
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-400 <PAR>
 A/Cross-references: GB:AL139075; GB:AL111168; NID:G6967817; PIDN:CAB75218.1; PID:G6968049
 A/Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:
 A/Gene: lysC; Cj0582
 C/Superfamily: aspartate kinase; aspartate kinase homology
 C/Keywords: phosphotransferase

Query Match 33.9%; Score 742.5; DB 2; Length 400;
 Best Local Similarity 39.9%; Pred. No. 1.4e-47;
 Matches 157; Conservative 89; Mismatches 110; Indels 37; Gaps 4;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 Db 4 VQYGGTSVGTLEIEIVANRIIATAASGKXVVVVVSAMSGVNTLIEQAEYFSKTPNGK 63
 Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTGTHGNAKIIDVTGRLQTA 120
 Db 64 DMDMLSSGERVTSALLALNEKGCFAISFGSKAGIITDSVFTKARLHHTDKTAIRSE 123
 Qy 121 LEEGRVVLVAGFGVQSDTKDVTTLGRGGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
 Db 124 LQNGKIVVIAGFGVDEE-GNVTTLGRGGSDLSAVAGALNADLCEIYTDVDFGVITDP 182
 Qy 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVWGS 240

```
Db      183 RIEPKAKKLDKISYEEMLESLGAKVLQNRSEVELAKKLNVLVTRSSFNNEGTMIT-- 240
Qy      241 IKDVPMDPILTVGAHRSKAKVTIVGLPDIPGYAAKVFAVARRRQHRHGAERLQGR 300
Db      241 -KEDMEQALVSGIALDKQARVTLRNIEDKPGIAAEIFSVLA----- 282
Qy      301 GQRDRHLLHLLPQTSQPPPKQW-----TRSETRSASTQLLYDDHIGKVS 346
Db      283 --NENINVDMIQNVGDGATNLGFTVPQNELELAKVAKQKILSSKTTIESDSAVLKVSI 340
Qy      347 IGACMRSHPGVATFCBALAAVGVNIELISTSE 379
Db      341 VGVGKSHSGVASKAFKALADEGINIGMISTSE 373

RESULT 15
B70399
A:Title: Aquifex aeolicus
A:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C:Accession: B70399
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70399
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-415 <AQF>
A:Cross-references: GB:AB000726; NID:g2983612; PIDN:AAC07182.1; PID:g2983613; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: lysC
C:Superfamily: aspartate kinase; aspartate kinase homology
F:3-400/Domain: aspartate kinase homology <DKI>

Query Match      33.5%; Score 734.5; DB 2; Length 415;
Best Local Similarity 42.4%; Pred. No. 5.9e-47;
Matches 164; Conservative 76; Mismatches 124; Indels 23; Gaps 6;

Qy      1 VQKYGGSVADAEIRRVAREIVATKQGNVVVVVSAAGDITDDLLDLAQOVCAPP 60
Db      5 VQKFGGTSVGTLEIEICAKKVIKALEKGYKPVVSSAMAGETDSLIELAKKVH 64
Qy      61 ELDMLLTAGERISNALVAMAIETSLCAHARSFTGSQAGVITTTGTHGNAKIIDVTPGR 120
Db      65 EMDTLLAVGEQKATAPANTINKLGYPAVSLCGWQVPIITDNVHTKARKIGISRLL 124
Qy      121 LEEGRVVLVAGFQVSGDQTKDVTTLGRGSDTTAVMAAALGADVCEIYTDVDPISAD 180
Db      125 IEEGYIPVAGFQGVTEW-EITTLGRGSDTTAVALLAALAKAD-CBINTDVPGVFTAD 182
Qy      181 RIVNARKLDVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSYSDRPGTVVVS 240
Db      183 RIVNARKIKKISYEEMLESLGAKVQIRSEFAKYNVIRHVRSTFEEEGTWIVP- 241
Qy      241 IKDVPMDPILTVGAHRSKAKVTIVGLPDIPGYAAKVFAVARRRQHRHGA----- 294
Db      242 -EDEVMEKVAVRGITVDTKTFTVVRVPDKPGIAAEIFKAL-----GDAHIVVDM 291
Qy      295 --ERLQGRGRODRHLLHLLPQTSQPPPKWNTSRSETSSAQQLLYDDHIGKVS 352
Db      292 IVQNVSHGEGYDMS--FTVNTKDADKAEIVKIAQKIGAEVVRDDKAKISIVGIGMK 349
Qy      353 SHPGVATFCBALAAVGVNIELISTSE 379
Db      350 SAYGVAGRMFDILARNGINIKAIISTSE 376
```

Search completed: November 21, 2003, 16:09:51
Job time : 9.84535 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 4.9955 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-4

Perfect score: 2193

Sequence: 1 VQYGGSSVADAEIRRVAE.....SAATRRPCTGRDGRWACQ 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match %	Length	DB ID	Description
1	1615.5	73.7	421	1	AK_MYCTU	P97048 mycobacteri
2	1478.5	67.4	421	1	AK_MYCSM	P41403 mycobacteri
3	1248.5	56.9	421	1	AK_COREF	Q8rqnl corynebacte
4	1224.5	55.8	421	1	AK_CORGL	P26512 corynebacte
5	1216.5	55.5	421	1	AK_CORFL	P41398 corynebacte
6	823	37.5	412	1	AK_PSEAE	O69077 pseudomonas
7	819	37.3	405	1	AK_THETH	P97151 thermus the
8	785	35.8	411	1	AK2_BACSP	O59228 bacillus sp
9	778	35.5	407	1	AK2_BACST	P53553 bacillus st
10	734.5	33.5	415	1	AK_AQUAE	O67221 aquifex aeo
11	721	32.9	405	1	AK_HELPJ	Q9zjz7 helicobacte
12	712.5	32.5	408	1	AK2_BACSU	P08495 bacillus su
13	711	32.4	405	1	AK_HELPY	O25827 helicobacte
14	532	24.3	404	1	AK1_BACSU	Q04795 bacillus su
15	492	22.4	473	1	AK1_MYCTU	Q57991 methanococc
16	480.5	21.9	446	1	AK1_RICRP	Q9zci7 rickettsia
17	364.5	16.6	819	1	AK1H_SERMA	P27725 serrattia ma
18	359.5	16.4	820	1	AK1H_ECOLI	P00561 escherichia
19	314.5	14.3	815	1	AKH_HAEIN	P44505 haemophilus
20	311.5	14.2	921	1	AKH_DAUCA	P37142 daucus caro
21	310.5	14.2	449	1	AK3_ECOLI	P08660 escherichia
22	306.5	14.0	454	1	AK3_BACSU	P94427 bacillus su
23	284	13.0	816	1	AKH_BUCAI	P54290 buchnera ap
24	280	12.8	431	1	AK1_CHLTR	O84367 chlamydia t
25	270	12.3	920	1	AKH1_MAIZE	P49079 zea mays (m
26	269	12.3	437	1	AK1_CHLMU	O9pk32 chlamydia m
27	266	12.1	917	1	AKH2_MAIZE	P49080 zea mays (m
28	264	12.0	814	1	AKH_EUCAP	Q8k9u9 buchnera ap
29	241	11.0	440	1	AK1_CHLPN	Q92610 chlamydia p
30	238	10.9	125	1	AK_MYCBO	P47731 mycobacteri
31	227.5	10.4	527	1	AK1_YEAST	P10869 saccharomyc
32	218	9.9	809	1	AK2H_ECOLI	P00562 escherichia
33	210	9.6	519	1	AK_SCHPO	O60163 schizosacch

ALIGNMENTS

RESULT 1

```

AK MYCTU          STANDARD;          PRT;   421 AA.
AC P97048: O69676; P97181;
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 15-SEP-2003 (Rel. 42, last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit (ASK-alpha); Aspartokinase beta subunit (ASK-beta)].
GN ASK OR RV3709C OR MT3812 OR MTV025.057C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RA Gilker J.M., Jucker M.T.;
RT "Mycobacterium tuberculosis ask-alpha, ask-beta and asd genes.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell S.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE (BY
CC SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
CC TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:

```

```

34 175 8.0 306 1 ARGB_STRCO Q11a3 streptomyc
35 174.5 8.0 239 1 PYRH_BARBA Q8rt65 bartonella
36 151 6.9 226 1 PYRH_METUA Q8656 methanococc
37 150 6.8 239 1 PYRH_FUSNN Q8r695 fusobacteri
38 149 6.8 318 1 ARGB_BIFLO P59295 bifidobacte
39 146 6.7 241 1 PYRH_AGRT5 Q8ufm1 agrobacteri
40 145.5 6.6 242 1 PYRH_ANASP Q8xxk5 anabaena sp
41 145.5 6.6 242 1 PYRH_ZYMMO Q8x5e9 zymomonas m
42 145 6.6 302 1 ARGB_STRCL Q8icse streptomyc
43 143.5 6.5 240 1 PYRH_XANCP P59009 xanthomonas
44 143 6.5 294 1 ARGB_MYCTU P49989 mycobacteri
45 141.5 6.5 240 1 PYRH_XANAC P59008 xanthomonas

```

CC Event=Alternative initiation;
 CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
 CC here) and Beta/Aspartokinase beta subunit, may be produced by
 CC alternative initiation;
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U90239; AAB49995.1; -;
 CC EMBL; U90239; AAB49994.1; -;
 CC EMBL; AL022121; CAA18031.1; -;
 CC EMBL; AB007178; AAK48180.1; -;
 CC PIR; F70794; F70794.
 CC TIGR; MT3812; -;
 CC TubercuList; Bv3709c; -;
 CC InterPro; IPR001048; Aa_kinase.
 CC InterPro; IPR002912; ACT.
 CC InterPro; IPR005260; Asp_kin_monofn.
 CC InterPro; IPR001341; Aspartate_kinase.
 CC Pfam; PF00696; aak_kinase; 1.
 CC Pfam; PF01842; ACT; 2.
 CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
 CC TIGRFAMs; TIGR00657; asp_kinases; 1.
 CC PROSITE; PS00324; ASPARTOKINASE; 1.
 CC Transferrase; Kinase; Diaminopimelate biosynthesis;
 CC Lysine biosynthesis; Alternative initiation; Complete proteome.
 CC CHAIN 1 421
 CC ASPARTOKINASE, ISOFORM ALPHA.
 CC FT CHAIN 250 250
 CC FOR ISOFORM BETA.
 CC FT INIT MET 250 250
 CC FT CONFLICT 288 336
 CC -----
 CC DADVNDVLQNVSKVEDGKTDITFTCSRDPVPAVSKLDS
 CC LTRNEIGFS -> RRRQRHGAERLQGRQGRDRLHLPLP
 CC QTSPPPPWNTSRSTRSAST (IN REF. 1).
 CC IRISVLCDTDLDAVALHEAFGLGDEEATVAVGTGR
 CC -> DORSCECAATNTRPWSRCKRSGSAAATRRPRCTRGR
 CC DGRWACO (IN REF. 1)
 CC SEQUENCE 421 AA; 44451 MW; 0567323BD02085A7 CRC64;
 CC -----
 CC Query Match 73.7%; Score 1615.5; DB 1; Length 421;
 CC Best Local Similarity 85.1%; Pred. No. 3e-106;
 CC Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;
 CC -----
 CC 1 VOKYGGSSVADAERIRRVAAERIVATKQGNVGVVSVAMGDTTDLDLAQOVCAPPPR 60
 CC 5 VOKYGGSSVADAERIRRVAAERIVATKQGNVGVVSVAMGDTTDLDLAQOVCAPPPR 64
 CC 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIDVTGRLQTA 120
 CC 65 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIDVTGRLQTA 124
 CC 121 LEEGKVLVAGFGQVSGQTKVITLGRGSGDTTAVAMAALGADVCIYTDVDFISADP 180
 CC 125 LEEGKVLVAGFGQVSGQTKVITLGRGSGDTTAVAMAALGADVCIYTDVDFISADP 184
 CC 181 RIVNRARKLDTVTPEEMLEMAACAKVLMRCVEYARRHNPVHVRSSYSDRPTVVVGS 240
 CC 185 RIVNRARKLDTVTPEEMLEMAACAKVLMRCVEYARRHNPVHVRSSYSDRPTVVVGS 244
 CC 241 IKDVPMPDILTVAHDRSEAKVTIVGLPDPGVAAKVFRVARRRRQHRHGAEE----- 295
 CC 245 IKDVPMPDILTVAHDRSEAKVTIVGLPDPGVAAKVFRVARRRRQHRHGAEE----- 295
 CC 296 RLQGRGRDRHLLHLLPOTS-----GPPPPWNTSRSTRSASTQLLYDDHIGKVLISLIGAMR 352
 CC 296 VLQNVSKVEDGKTDITFTCSRDPVPAVSKLDSLENEIGFSQLLYDDHIGKVLISLIGAMR 355
 CC 353 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392
 AK_MYCSM STANDARD; PRT; 421 AA.
 ID AC MYC3M
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
 DE alpha subunit (ASK-alpha); Aspartokinase beta subunit (ASK-beta)].
 DE ASK.
 GN Mycobacterium smegmatis.
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 607 / mc(2)6 / NRRL B-692;
 RX MEDLINE=94254720; PubMed=7910936;
 RA Cirillo J.D., Weisbrod T.R., Pascopella L., Bloom B.R.,
 RA Jacobs W.R. Jr.;
 RT "Isolation and characterization of the aspartokinase and aspartate
 RT semialdehyde dehydrogenase operon from mycobacteria.";
 RL Mol. Microbiol. 11:629-639(1994).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 CC aspartate.
 CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE (BY
 CC SIMILARITY).
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 CC FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
 CC TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
 CC isoforms Beta (function not known) (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
 CC here) and Beta/Aspartokinase beta subunit, may be produced by
 CC alternative initiation;
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z17372; CAA78984.1; -;
 CC EMBL; Z17372; CAA78985.1; ALT_INIT.
 CC PIR; S42422; S42422.
 CC InterPro; IPR001048; Aa_kinase.
 CC InterPro; IPR002912; ACT.
 CC InterPro; IPR005260; Asp_kin_monofn.
 CC InterPro; IPR001341; Aspartate_kinase.
 CC Pfam; PF00696; aak_kinase; 1.
 CC Pfam; PF01842; ACT; 2.
 CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
 CC TIGRFAMs; TIGR00657; asp_kinases; 1.
 CC PROSITE; PS00324; ASPARTOKINASE; 1.
 CC Transferrase; Kinase; Diaminopimelate biosynthesis;
 CC Lysine biosynthesis; Alternative initiation.
 CC CHAIN 1 421
 CC ASPARTOKINASE, ISOFORM ALPHA.
 CC FT CHAIN 250 421
 CC ASPARTOKINASE, ISOFORM BETA.
 CC FT INIT MET 250 250
 CC FOR ISOFORM BETA.
 CC SEQUENCE 421 AA; 44458 MW; 92689FC8625E20CD CRC64;
 CC -----
 CC Query Match 67.4%; Score 1478.5; DB 1; Length 421;
 CC Best Local Similarity 77.6%; Pred. No. 1.2e-96;
 CC Matches 302; Conservative 31; Mismatches 35; Indels 21; Gaps 4;


```
Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VQYGGSSVADAEIRIRVAERIVETKAGNDVVVVSAMGDTTDDLLDLARQVSPAPPR 64
Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 124
Qy 121 LEEGRVVLVAGFGQVGSODTKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVDFPSADP 180
Db 125 LDEGQIVLVAGFGQVGSQSDKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVDFPSADP 184
Qy 181 RIVNARKLDTVTFFEMLENAACGAKVLMRCVEYARRHNI PVHVRSSYSYSDRGTVVGS 240
Db 185 RIVNARHLDTVTFEEMLENAACGAKVLMRCVEYARRHNI PVHVRSSYSYSDRGTVVGS 244
Qy 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFAVARRRQHRHGAAE----- 295
Db 245 IEDIPMEDAILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFAVARRRQHRHGAAE----- 295
Qy 296 RLQG-----RGRQRHLLHLLPQTSQPPKWNTRSETRSASTQLLYDDHIGKVSILGAG 350
Db 296 VLQNIKIEGDKTD--ITFTCARDNGPRAVEKLSALKSEIGFQSVLYDDHIGKVSILGAG 353
Qy 351 MRSPGVTATFCEALAAVGVNIELISTSE 379
Db 354 MRSPGVTATFCEALAAEAGINIDLISTSE 382

RESULT 3
AK COREP
ID AK COREP STANDARD; PRT; 421 AA.
AC OSEQL;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit; Aspartokinase beta subunit].
GN LVSC OR ASK OR CE0220.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN SEQUENCE FROM N.A.
RP STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RC Itaya H., Kimura E., Kawahara Y., Hino Y., Kikuchi H., Nakamura Y.,
RA Kawaiabavasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO THE CELL WALL PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
CC TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms (Alpha (catalytic) and 2
CC isoforms Beta (function not known)).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation;
CC Comment-2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, are produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB081129; BAB88820.1; -; ALT_INIT.
DR EMBL; AF005214; BAC17030.1; -; ALT_INIT.
DR InterPro; IPR001048; AA_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakkinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfam; TIGR00656; asp_kin_monofn; 1.
DR TIGRfam; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Alternative initiation; Complete proteome.
FT CHAIN 1 421 ASPARTOKINASE, ISOFORM ALPHA.
FT CHAIN 250 421 ASPARTOKINASE, ISOFORM BETA.
FT INIT_MET 250 250 FOR ISOFORM BETA.
SQ SEQUENCE 421 AA; 44793 MW; F0D638306983CB5F CRC64;

Query Match 56.9%; Score 1248.5; DB 1; Length 421;
Best Local Similarity 65.2%; Pred. No. 1.7e-80;
Matches 251; Conservative 58; Mismatches 63; Indels 13; Gaps 3;

Qy 1 VQYGGSSVADAEIRIRVAERIVATKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VQYGGSSLESARIRVAERIVATKAGNDVVVVSAMGDTTDELLDLAAVNPVPPAR 64
Qy 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
Db 65 EMDMLLTAGERISNALVAMAEISLGAARSFTGSQAGVITTTGHNARIVDVTGVRREA 124
Qy 121 LEEGRVVLVAGFGQVGSQDKVTTLGRGGSDTTAVAMAAALGADVCEIYTDVDFPSADP 180
Db 125 LDEGKICIVAGFGQVGNKTRDVTTLGRGGSDTTAVAMAAALGADVCEIYSDVGVYADP 184
Qy 181 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSYSDRGTVVGS 240
Db 185 RIVNARQLERLSFEEMLEAAVGSKILVRSVEYARRHNI PVHVRSSYSYSDRGTVVGS 244
Qy 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDIPGYAAKVFAVARRR-----RQHRHGAA 294
Db 245 MEDIPWEBAVLGTGATDKSEAKVTIVGLIPDKGCEAAKVFRALADAEINIDMVLVNVSSVE 304
Qy 295 ERLQGRGRDRHLLHLLPQTSQPPKWNTRSETRSASTQLLYDDHIGKVSILGAGMRSH 354
Db 305 D-----GTTD--ITFTCPRSDDGPRAVELLKRMQGGDWTNVLVYDDQGVKSLVAGMKSH 357
Qy 355 PGVTATFCEALAAVGVNIELISTSE 379
Db 358 PGVTATFCEALAAVGVNIELISTSE 382

RESULT 4
AK COREP
ID AK COREP STANDARD; PRT; 421 AA.
AC P76512; Q59286;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit; Aspartokinase beta subunit].
GN LVSC OR CGL0251.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
```

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RX MEDLINE=92065816; PubMed=1956296;
 RA Kalinowski J., Cremer J., Bachmann B., Eggeling L., Sahm H.,
 RA Puehler A.;
 RT "Genetic and biochemical analysis of the aspartokinase from
 RT Corynebacterium glutamicum";
 RL Mol. Microbiol. 5:1197-1204 (1991).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RC Nakagawa S.;
 RA "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 1-51 FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RC MEDLINE=94161495; PubMed=8117072;
 RA Patek M., Krumbach K., Eggeling L., Sahm H.;
 RT "Leucine synthesis in Corynebacterium glutamicum: enzyme activities,
 RT structure of leuA, and effect of leuA inactivation on lysine
 RT synthesis";
 RL Appl. Environ. Microbiol. 60:133-140 (1994).
 RN [4]
 RN SEQUENCE OF 158-421 FROM N.A.
 RP STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RC MEDLINE=91094767; PubMed=1980002;
 RA Kalinowski J., Bachmann B., Thierbach G., Puehler A.;
 RT "Aspartokinase genes lysC alpha and lysC beta overlap and are
 RT adjacent to the aspartate beta-semialdehyde dehydrogenase gene asd in
 RT Corynebacterium glutamicum";
 RL Mol. Genet. 224:317-324 (1990).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 CC aspartate.
 CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY LYSINE AND THREONINE.
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 CC FROM ASP TO THE CELL WALL. PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
 CC TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: Tetramer consisting of two isoforms Alpha (catalytic) and
 CC two isoforms Beta (function not known).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative initiation;
 CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
 CC here) and Beta/Aspartokinase beta subunit, are produced by
 CC alternative initiation;
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57226; CAA40502.1; -
 DR EMBL; X57226; CAA40503.1; -
 DR EMBL; AF005274; BAB97644.1; -
 DR EMBL; X70959; CAA50296.1; ALT_SEQ.
 DR PIR; I40723; I40723
 DR PIR; S15276; S15276
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR InterPro; IPR001341; Aspartate_kinase.
 DR Pfam; PF00696; askinase; 1.
 DR Pfam; PF01842; ACT; 2.
 DR TIGRfams; TIGR00656; asp_kin_monofn; 1.
 DR TIGRfams; TIGR00857; asp_kinases; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 DR Transferrase; Kinase; Diaminopimelate biosynthesis;
 KW Lysine biosynthesis; Alternative initiation; Complete proteome.
 CC Lysine biosynthesis; 1
 CC 421 ASPARTOKINASE, ISOFORM ALPHA.
 CC CHAIN

FT CHAIN 250 421 ASPARTOKINASE, ISOFORM BETA.
 FT INIT MET 250 250 FOR ISOFORM BETA.
 FT MUTAGEN 301 301 S->Y: FEEDBACK-RESISTANT AND ENHANCED
 FT EXPRESSION OF THE ASD GENE.
 FT CONFLICT 40 40 C -> V (IN REF. 1 AND 3).
 SQ SEQUENCE 421 AA; 44754 MW; E36B4D0081DE0827 CRC64;
 Query Match 55.8%; Score 1224.5; DB 1; Length 421;
 Best Local Similarity 64.0%; Pred. No. 8.1e-79;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VQYGGSSVADAEIRIRVVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPPR 60
 DB 5 VQYGGSSLESERIRNVAERIVATKKAGNDVVVVC SAMGDTTDELELAAAANVPVPPAR 64
 QY 61 ELDMLLTAGERISNALVAMAEISLGAHARSFTQSQAGVITTTGTHGNAKIIDVTGRLQTA 120
 DB 65 EMDYLLTAGERISNALVAMAEISLGAHARSFTQSQAGVITTTGTHGNARIVDTVPGRVRA 124
 QY 121 LLEGVVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAAALGADYCEIYTDVDFPSADP 180
 DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSDTTAVAAALANADYCEIYSDVDGVYADP 184
 QY 181 RIVNARKLDTVTFEEMLEMAACGKVLMLRCVEYARRHNI PVHVRSSYSDRPGTVVCS 240
 DB 185 RIVNQAQKLEKSFEEMLLAAGVSKILVRSVEYARAFNVPLVRSSYSDNPGLTIAGS 244
 QY 241 IKQVPMEDPILTGVAHDSERAKVTIVGLPDIIPGYAAKVFRAVA-----RRRRQHR 290
 DB 245 MEDIPVEEAIVTGATDKSEAKVTVLGISDKPCEAAKVFALADAEINIDWLQNVSSVE 304
 QY 291 HGAERLQGRGROD-RHLLHLLP--QTSGPPKWKWTRSETRSASTOLLYDHDHIGKVSLL 347
 DB 305 DGTDTITFTCPSSDGRRAEILKKLVQVG-----TNVLYDDQVGKVSIV 350
 QY 348 GAGMRSHPGVTATFCALAAVGVNIELISTSE 379
 DB 351 GAGMKSHPGVTAEFFMALRDVNVNIELISTSE 382
 RESULT 5
 AK CORFL STANDARD; PRT; 421 AA.
 ID AK CORFL
 AC P41398;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
 DE alpha subunit; Aspartokinase beta subunit].
 GN LYSC OR ASK.
 OS Corynebacterium flavam.
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriidae; Corynebacterium.
 OX NCBI_TaxID=28028;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=N13;
 RX MEDLINE=93308089; PubMed=8100567;
 RA Follettie M.T., Peoples O.P., Agoropoulou C., Sinskey A.J.;
 RT "Gene structure and expression of the Corynebacterium flavam N13
 RT ask-asd operon";
 RL J. Bacteriol. 175:4096-4103 (1993).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 CC aspartate.
 CC -!- ENZYME REGULATION: LOW FEEDBACK INHIBITION BY LYSINE AND
 CC THREONINE. THE DECREASED SENSITIVITY OF C FLAVUM, AS COMPARED TO
 CC C. GLUTAMICUM, IS PROBABLY DUE TO THE D345G SEQUENCE DIFFERENCE.
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 CC FROM ASP TO THE CELL WALL. PRECURSOR MESO-DIAMINOPIMELATE, TO LYS,
 CC TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
 CC isoforms Beta (function not known) (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:

```
CC Event=Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, are produced by
CC alternative initiation;
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16848; AAA23293.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
FT CHAIN 1 421
FT INIT 250 421
FT INIT_MET 250 250
FT MUTAGEN 345 345
FT ACTIVITY TO CONCERTED FEEDBACK INHIBITION
FT BY LYSINE AND THREONINE.
FT
SQ SEQUENCE 421 AA; 44796 MW; 4B37B0B6597F665E CRC64;
Query Match 55.5%; Score 1216.5; DB 1; Length 421;
Best Local Similarity 63.8%; Pred. No. 3e-78;
Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
QY 1 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVVSAMGDTTDDLLDLAQVCPPAPPPR 60
DB 5 VQKYGSSLSAEIRIRVAERIVATKKAGNDVVVVCVSAMGDTTDELLELAANVPVPPAR 64
QY 61 ELDMLLTAGERISNALVAMATESIAGHARSFTGSCAGVITTTGTHGNAKIIDVTPGRLQTA 120
DB 65 EMDMLLTAGERISNALVAMATESIAGAEQFTGQAGVLITERHGNARIVDVTPGVRREA 124
QY 121 LEEGRVVLVAGFQGVSDQTDVTTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 180
DB 125 LDEKICIVAGFQGVNKETRDVTTTLGRGSGDTTAVALLANADVCIIYSDVDGVYTDAP 184
QY 181 RIVNARKLDTVTVEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRGTVTVVGS 240
DB 185 RIVNPAQKLEKLSPEEMLELAAGSKILVLSRVEYARAFNVPLFVRSSYNDPGLIAGS 244
QY 241 IKDVPMDPILTVGAHDSRSEAKVTIGVLPDIPGVAAKVFRAVA-----RRRQHR 290
DB 245 MEDIPVEAVLTGATDKSEAKVTIGLISDYPGEAKVFRALADABINIDMVLQNVSSE 304
QY 291 HGAERLQGRQD-RHHLHLLP--QTSPPPPKWNWTRSETRSASTOLLDDHIGKVSLLI 347
DB 305 DGTDDITFTCPADGRRAAMEIKKLQVG-----NW-----TNVLYDDQDKVSLV 350
QY 348 GAGMRSHPGVTATCEALAAVGVNIELISTSE 379
DB 351 GAGMRSHPGVTAEFMEALRDVNVNIELISTSE 382
RESULT 6
AK_PSEAE STANDARD; PRT; 412 AA.
AC O63077; Q91552;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR PA0904.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [2]
RP SEQUENCE OF 198-412 FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Hindle Z., Throup J.P., Francis K.P., Bycroft B.W., Williams P.,
RA Stewart G.S.A.B.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -1- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF004525; AAG04293.1; -.
DR EMBL; AF061757; AAC16241.1; -.
DR PIR; C83531; C83531.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
SQ SEQUENCE 412 AA; 44388 MW; A06ACAA5E914F337 CRC64;
Query Match 37.5%; Score 823; DB 1; Length 412;
Best Local Similarity 42.6%; Pred. No. 1.2e-50;
Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;
QY 1 VQKYGSSVADAEIRIRVAERIVATKKQNDVVVVVVSAMGDTTDDLLDLAQVCPPAPPPR 60
DB 5 VQKFGGTSVGTVERIEQVAEKVKFKREAGDDVVVVVSAMSGETNRLIGLANQIMEQVPR 64
QY 61 ELDMLLTAGERISNALVAMATESIAGHARSFTGSCAGVITTTGTHGNAKIIDVTPGRLQTA 120
DB 65 ELDVWVSTGEQVTIALLSNALIKRGVPVSVYTGNGVNRILTDSAKTKARLHIDDTTHRAD 124
QY 121 LEEGRVVLVAGFQGVSDQTDVTTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 180
DB 125 LKAGRVVVVAGFQGV-DGNGNITTLGRGSGDTTGVALLAALKADECIYTDVGVYTTDP 183
QY 181 RIVNARKLDTVTVEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRGTVTVVGS 240
```

```

Db 184 RVVPQARRLDKITFEEMLEMASLGSVQLQIRAVEFAGKYNVPLRLVHLSFOEGFGTLITD 243
QY 241 IKDVPMDPLITGVADRSEAKVTIVGLPDIPGYAAKVFRAVARRRQ-----288
Db 244 DEESMEQQLISIAINRDEAKLTIRGVPDTPGVAFKILGPISAANVEVDMTVQVADHN 303
QY 289 -----HRHGAERLOGRQDRHLLHLLPQTSGPPPKWNTSRSEASTQLLYDDH 340
Db 304 TTDFTVHRNDYLALE-----ILKQTA-----ANIGAREAIGDTN 340
QY 341 IGKVSILIGMSPHQVTFATFCEALAAVGVNIELISTSE 379
Db 341 IAKVSIVGVGRSHAGVAGRMFEALAKESINQIMISTSE 379

RESULT 7
AK_THETH STANDARD; PRT; 405 AA.
AC P97151; P77991;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase) [Contains: Aspartokinase
DE alpha subunit (ASK-alpha); Aspartokinase beta subunit (ASK-beta)].
GN ASK OR ASKAB.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=99173891; PubMed=10074061;
RA Kobashi N., Nishiyama M., Tanokura M.;
RA Nishiyama M., Kukimoto M., Beppu T., Horiuchi S.;
RT "An operon encoding aspartokinase and purine phosphoribosyltransferase
RT in Thermus flavus";
RL J. Bacteriol. 181:1713-1718(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33923 / DSM 674 / AT-62;
RX MEDLINE=95291465; PubMed=7773416;
RA Nishiyama M., Kukimoto M., Beppu T., Horiuchi S.;
RT "An operon encoding aspartokinase and purine phosphoribosyltransferase
RT in Thermus flavus";
RL Microbiology 141:1211-1219(1995).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase alpha subunit (shown
CC here) and Beta/Aspartokinase beta subunit, may be produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB013131; BAA25848.1; -.
CC EMBL; AB013131; BAA25849.1; -.
CC EMBL; D37928; BAA07146.1; -.
CC EMBL; D37928; BAA07147.1; -.
CC InterPro; IPR001048; Aa_kinase.

```

```

DR InterPro; IPR002912; ACT.
DR InterPro; IPR005260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMS; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Alternative initiation.
FT CHAIN 1 405 ASPARTOKINASE, ISOFORM ALPHA.
FT CHAIN 245 405 ASPARTOKINASE, ISOFORM BETA.
FT INIT MET 245 245 FOR ISOFORM BETA.
FT VARIANT 126 126 E -> D (IN STRAIN AT-62).
SQ SEQUENCE 405 AA; 43333 MW; 06F7B79E908960EA CRC64;

Query Match 37.3%; Score 819; DB 1; Length 405;
Best Local Similarity 43.2%; Pred. NO. 2.2e-50;
Matches 175; Conservative 78; Mismatches 94; Indels 58; Gaps 6;

QY 1 VKYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDLDDLAQVCPAPPR 60
Db 5 VQYGGTSVGDLEIRIHKVAQRIAHYEKEGHRLLAVVVSAMGHTTDLIALAKVNRPPR 64
QY 61 ELDMLTAGERISNALVMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTGRLQTA 120
Db 65 ELDLTTTGEQSVALLSMQWAMGIPAKGFVQHQTGITTDGRYGDARILEVNPAREA 124
QY 121 LEEGRVVLVAGFOGVSQDTKVTTLGRGSDTTAVAMAALGADVCEIYTDVDGIFSADP 180
Db 125 LEQGFVAVIAGFMGTPE-GEITTLGRGSDTTAVAAALGAKECEIYTDIEGYTTDP 183
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHRSSYSRDEPGTVVGS 240
Db 184 HLIPKARKLSVIGYDQMLEWALGARVLPRAVYAKRYGVVLHVRSSPSYNGTLV--- 240
QY 241 IKDVPME-DPLITGVADRSEAKVTIVGLPDIPGYAAKVFRAVARR-----285
Db 241 -KEVAMEMKAVTGVVALDLCHQIGLIGIPDOPGIAKVFQALAEAGIADVMIQGVPH 299
QY 286 -----RQHRHGAERLOGRQDRHLLHLLPQTSGPPPKWNTSRSEASTQ 334
Db 300 DPSPQQWAFVTKDPAQEALE-----PVLAEIGG-----E 332
QY 335 LLYDDHIGKVLIGAGMRSHPGVTFATFCEALAAVGVNIELISTSE 379
Db 333 AILRPDIKVSIVGVGLASTPEVPAPKMFQAVASTGANIEMIATSE 377

RESULT 8
AK2_BACSP STANDARD; PRT; 411 AA.
AC Q99223;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta
DE subunit].
DE LysC.
GN Bacillus sp. (strain MGA3).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93073880; PubMed=1444390;
RA Schendel F.J., Flickinger M.C.;
RT "Cloning and nucleotide sequence of the gene coding for aspartokinase
RT II from a thermophilic methylophilic Bacillus sp.";
CC Appl. Environ. Microbiol. 58:2806-2814(1992).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE.

```

```

CC CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC CC isoforms Beta (function not known).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event-Alternative initiation;
CC CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown
CC CC here) and Beta/Aspartokinase II beta subunit, are produced by
CC CC alternative initiation;
CC CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M93419; AAA22251.1; -.
CC CC InterPro: IPR001048; Aa_kinase.
CC CC InterPro: IPR002912; ACT.
CC CC InterPro: IPR005260; Asp_kin_monofn.
CC CC InterPro: IPR001341; Aspartate_kinase.
CC CC Pfam; PF00696; aakinas; 1.
CC CC Pfam; PF01842; ACT; 2.
CC CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC CC PROSITE; PS00324; ASPARTOKINASE; 1.
CC CC Transferrase; Kinase; Diaminopimelate biosynthesis;
CC CC Lysine biosynthesis; Alternative initiation.
CC CC CHAIN 1 411 ASPARTOKINASE 2, ISOFORM ALPHA.
CC CC FT CHAIN 246 411 ASPARTOKINASE 2, ISOFORM BETA.
CC CC FT INIT MET 246 246 FOR ISOFORM BETA.
CC CC SQ SEQUENCE 411 AA; 44342 MW; 94061572E369CD48 CRC64;

Query Match 35.8%; Score 785; DB 1; Length 411;
Best Local Similarity 42.1%; Pred. No. 5.5e-48;
Matches 170; Conservative 81; Mismatches 97; Indels 56; Gaps 6;

Qy 1 VQKYGSSVADAEIRIRVAERIVATKKQGNVVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VQKFGTSGVSVIRILVANNRIVEEKKNGNDVVVVVSAMGKTTDELVDLAKQISAPPKR 64

Qy 61 ELDMLITAGERISNALVAMAIESLGAHARSFTGSGAGVITTGTHGNAKIIDVTPGLQTA 120
Db 65 EMDMLITGEGVITSLALNALNEKGYEAISTGQAGITTEPFGNARILNIETEKIQK 124

Qy 121 LEEGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFADP 180
Db 125 LNEGKIVVVGFGQIDEH-GEITTLGRGSGDTTAVALLAALKAECIDIYDVTGVFTDP 183

Qy 181 RIVNARKLDVTVEEMLEMAACAGKVLRCVYARRHNPVHVRSGSYSDRPGTVVGS 240
Db 184 RYVSARKLASISYDEMLELANLGAVLHPRAVEFAKNYGTITLVRSSMEREEGTIE-- 241

Qy 241 IKQVPE--DPILTGVADHRSKATVIGLPIPGVAAKRVFRAVRRRRQRHGAERLQ 299
Db 242 -EVTMEQNLVVRGVAFDEITRVTFGLPNSLSLSTIFTTLA----- 284

Qy 300 RGRDRHLLHPQTSPPPPKNNTRSTRASATQLLYDD----- 339
Db 285 ---QNRINVDIIQSA-----TDAETTLFSIKSDDLLETMAVLENNKNLNYQGI 333

Qy 340 ---HIGKVSILGAGMRSPGVATFCEALAAVGNVLIETSE 379
Db 334 BSETGLAKVISVGSNINPVAAKMFEVLALNGIQVKNVTS 377

RESULT 9
AK2_BACST
ID_AK2_BACST
AC_P53553;

```

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta
DE subunit].
GN LysC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANUB3621;
RX MEDLINE=96186919; PubMed=8635739;
RA Cantoni R., Labo M., de Rossi E., Riccardi G.;
RT "Sequence of the Bacillus stearothermophilus gene encoding
RT aspartokinase II.";
RL Gene 169:135-136(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown
CC here) and Beta/Aspartokinase II beta subunit, may be produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; J46351; AAB06216.1; -.
CC CC FIR; J4640; JC4640.
CC CC InterPro: IPR001048; Aa_kinase.
CC CC InterPro: IPR002912; ACT.
CC CC InterPro: IPR005260; Asp_kin_monofn.
CC CC InterPro: IPR001341; Aspartate_kinase.
CC CC Pfam; PF00696; aakinas; 1.
CC CC Pfam; PF01842; ACT; 2.
CC CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC CC PROSITE; PS00324; ASPARTOKINASE; 1.
CC CC Transferrase; Kinase; Diaminopimelate biosynthesis;
CC CC Lysine biosynthesis; Alternative initiation.
CC CC CHAIN 1 407 ASPARTOKINASE 2, ISOFORM ALPHA.
CC CC FT CHAIN 246 407 ASPARTOKINASE 2, ISOFORM BETA
CC CC (PROBABLE).
CC CC FT INIT MET 246 246 FOR ISOFORM BETA (PROBABLE).
CC CC SQ SEQUENCE 407 AA; 44002 MW; 2A848E6A7037B172 CRC64;

Query Match 35.5%; Score 778; DB 1; Length 407;
Best Local Similarity 43.2%; Pred. No. 1.7e-47;
Matches 173; Conservative 77; Mismatches 98; Indels 52; Gaps 7;

Qy 1 VQKYGSSVADAEIRIRVAERIVATKKQGNVVVVVSAMGDTTDDLDAQQVCPPPPR 60
Db 5 VQKFGTSGVSVIRILVANNRIVEEKKNGNDVVVVVSAMGKTTDELVDLAKQISNPKR 64

Qy 61 ELDMLITAGERISNALVAMAIESLGAHARSFTGSGAGVITTGTHGNAKIIDVTPGLQTA 120
Db 65 EMDMLITGEGVITSLALNALNEKGYKAVSLTGWQAGITTEEMHGNARINIDTTIRRC 124

Qy 121 LEEGRVVLVAGFQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFADP 180

```

```

Db 125 LDEGAIVIVAGFGVGT-ETGEITTLGRGGSDTTAVALAAALKAEKCDIYDVTGVTFTDP 183
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVVVG 240
Db 184 RYKTKARKIEISYDEMLELANLGAGVLPRAVEFAKNVEPLEVRSSMENERGTMRK- 241
QY 241 IKDVPMEDEP-ILITGVADHRSSEAKVTIVGLPDIPGVAAKVFRVARRRRQHRHGAERLQG 299
Db 242 -BEVSEQHILVIRGIAFEQDVTRVTVGIEKVLQSVARTIFALANR----- 286
QY 300 RGRQDRHLLHLLPQTSQPPPKWNTSETRSAS-----TOLL-----YDD 339
Db 287 -----GINDIITIQ-----NATNSETASVSFSTRTDLPETLQVLQALEGADVHVES 333
QY 340 HIGKVLGAGMRSHPGVTATCEALAAVGVNIELISTSE 379
Db 334 GLAKSVISVSGMISNPGVAARVFEVLADQGIKMWISSE 373

RESULT 10
AK_AQUAE STANDARD; PRT; 415 AA.
AC O67221;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR AQ.1152.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.:
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358 (1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000726; AAC07182.1; --
CC PIR; B70399; B70399.
CC InterPro; IPR001048; Aa kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR005260; Asp_kin_monofn.
CC InterPro; IPR001341; Aspartate_kinase.
CC Pfam; PF00696; aak kinase; 1.
CC Pfam; PF01842; ACT; 2.
CC TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
CC TIGRFAMs; TIGR00657; asp_kinases; 1.
CC PROSITE; PS00324; ASPARTOKINASE; 1.
CC Transferrase; Kinase; Diaminopimelate biosynthesis;
CC Lysine biosynthesis; Complete proteome.
CC SEQUENCE 415 AA; 45519 MW; E811750A8C14162 CRC64;

Query Match 33.5%; Score 734.5; DB 1; Length 415;
Best Local Similarity 42.4%; Pred. No. 1.9e-44;
Matches 164; Conservative 76; Mismatches 124; Indels 23; Gaps 6;

```

```

QY 1 VQYGGSSVADAEIRIRRVATKQKQNDVVVVVSAMGDTTDDLLDLAQVCPAPPR 60
Db 5 VQKFGTSTVGLTERIELCAKRVIKALEKGYKPVVSSAMAGETDSLIELAKKVHPPPR 64
QY 61 ELDMLTAGERINALVAMAEISLGAHARSFTQSGAVITTGTHGNAKIIDVTPGRLQTA 120
Db 65 EMTLLAVEGEKALAFMTNLKLGVPVSLCWQVPIITDNTVHTKARKIKISLLSL 124
QY 121 LEBGRVVLVAGFGVQSDTKQVTTLGRGGSDTTAVAMAALGADVCEIYTDVDFGSADP 180
Db 125 IEERYIPVWAGFGVQVTEDEW-EITTLGRGGSDTTAVALLAALRAD-CEINTDVGFTADP 182
QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVVVG 240
Db 183 RIVNARKIKKISYEEMLELASLGAKVMQIRSVFAAKYNVRIHVRSTFESEEGTWIVP- 241
QY 241 IKDVPMEDEPILITGVADHRSSEAKVTIVGLPDIPGVAAKVFRVARRRRQHRHGA- 294
Db 242 -EDEVMEKVAARGITVDTKTETFTVVRVDPKPGIAAEIFKAL-----GDAHIVVDM 291
QY 295 --RELQGRGRDRHLLHLLPQTSQPPPKWNTSETRSASQTLLYDDHIGKVLGAGMR 352
Db 292 IVQNVSHGYTDM--FTVNTKDADKABEIVKIKAQIKGAEVREDKVKAKISIVIGMK 349
QY 353 SHPGVTATCEALAAVGVNIELISTSE 379
Db 350 SAYGVAGRMFDILARNGINIKALISTSE 376

RESULT 11
AK_HELPU STANDARD; PRT; 405 AA.
AC Q9JZ7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN LYSC OR JHP1150.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Urias-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180 (1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE, TO MET, TO ILE AND TO THR.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001542; AAD06723.1; --
CC PIR; H71843; H71843.
CC InterPro; IPR001048; Aa kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR005260; Asp_kin_monofn.

```

```
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00596; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfam; TIGR00656; asp_kin_monofn; 1.
DR TIGRfam; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
DR TRANSFERASE; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
SQ SEQUENCE 405 AA; 44219 MW; E43A0D62E890A8EF CRC64;

Query Match 32.9%; Score 721; DB 1; Length 405;
Best Local Similarity 41.4%; Pred. NO. 1.7e-43;
Matches 161; Conservative 84; Mismatches 120; Indels 24; Gaps 6;

QY 1 VQYGGSSVADAEIRIRVRAIVATKQGNVWVVSAMGDTTDLDDLAQVCVAPP 60
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 VQYGTSGMSIERIHVAQVRVLESVKLGQVWVVSAMSGETDLRLEFGKFNPNKR 63
QY 61 ELMILTAGERISNALVAVAIISLGAHARSFTGSGQAGVTTGTHGNAKIIVTPORLQTA 120
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 EMDRIVSAGEWISSAALSALERYGHRALSLSGKEAGILTSHFQNAVQSIDTQRI 123
QY 121 LEEGRVIVAGFGVGSODTKDVTLLGRGSDTTAVAMAALGADVCEIYTDVDFGFSADP 180
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 LEKNYIVVIAGFGQ-ADIQETITLLGRGSDLSVALAGALKALHCEIYTDVGVYITDP 182
QY 181 RIVRNARKLDTVTFEEMLEMAACCAKVLMLRCVYARHNPVHVSSYSDRPGTVVVG- 239
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 RIBEKAQKIAQISYDEMLELASMCAKVLNRSVELAKLSVKLVTFNSFNHSEGLIVAE 242
QY 240 -STKVPMDPILTVGAHDBSEAKVTIVGLPDIPGVAAKVRAVARRRQHSHGAERLQ 298
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 KDFKGERMETPIVSGIALDKQARVMEGVDEDPGIAAEIFGALA-----EYINVDIMVQ 298
QY 299 GRGQRDRHLEH-LPQTSQPPPPKQWTRSTRSASTQLL-----YDDHICKVSLIGAG 350
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 TIGRDKGKTDLDFTIVKQ-----IEETKALKPFLAQMDSIDYDENIAKVISVGVG 349
QY 351 MRHPGVATFCBALAAVGNIELISTSE 379
DB 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 MKSHSGVASIAFKALADKNINIMISTSE 378

RESULT 12
AK2_BACSU STANDARD; PRT; 408 AA.
AC P08495; P08496; P94554; P97183;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartokinase 2 (EC 2.7.2.4) (Aspartokinase II) (Aspartate kinase 2)
DE [Contains: Aspartokinase II alpha subunit; Aspartokinase II beta subunit].
GN LVSC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250503; PubMed=3036830;
RA Chen N.-Y., Hu F.M., Paulus H.;
RT "Nucleotide sequence of the overlapping genes for the subunits of
RT Bacillus subtilis aspartokinase II and their control regions.";
RL J. Biol. Chem. 262:8787-8798 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;
RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emmerson P.T., Harwood C.R.;
RT "The dnaB-phaA (256 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress responses,
RT the utilization of plant cell walls and primary metabolism.";
RL J. Biol. Chem. 262:8787-8798 (1987).

Microbiology 142:3067-3078 (1996).
[3]
RN SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Enrian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi E., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha B., Roche B., Rose M., Sadale Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serfor P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambat R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RP ALTERNATIVE INITIATION.
RX MEDLINE=98243847; PubMed=2837491;
RA Chen N.-Y., Paulus H.;
RT "Mechanism of expression of the overlapping genes of Bacillus
RT subtilis aspartokinase II.";
RL J. Biol. Chem. 263:9526-9532 (1988).
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -!- ENZYME REGULATION: LYSINE-SENSITIVE.
CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -!- SUBUNIT: Tetramer consisting of 2 isoforms Alpha (catalytic) and 2
CC isoforms Beta (function not known).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, Alpha/Aspartokinase II alpha subunit (shown
CC here) and Beta/Aspartokinase II beta subunit, are produced by
CC alternative initiation;
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03294; AAA87318.1; -
CC EMBL; J03294; AAA87319.1; -
CC EMBL; Z75208; CAA9580.1; -
CC EMBL; Z75208; CAA9581.1; -
CC EMBL; Z99118; CAB14807.1; -
CC PIR; A29314; A29314.
CC Subtilist; BG10350; lysc.
CC InterPro; IPR01048; Aa_kinase.
```


30-MAY-2000 (Rel. 39, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartokinase 1 (EC 2.7.2.4) (Aspartokinase 1)
[Contains: Aspartokinase I alpha subunit; Aspartokinase I beta subunit].
DE DAPG OR LSSP.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93252813; PubMed=8098035;
RA Chen N.-Y., Jiang S.-Q., Klein D.A., Paulus H.,
RT "Organization and nucleotide sequence of the Bacillus subtilis
RT diaminopimelate operon, a cluster of genes encoding the first three
RT enzymes of diaminopimelate synthesis and dipicolinate synthase.",
RL J. Biol. Chem. 268:9448-9465(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Arevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschini C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kaeahara Y., Klaert-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,
RA Kurita K., Lapicque A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche E., Rose B., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Teateo V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzensegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -1- ENZYME REGULATION: DIAMINOPIMELATE-SENSITIVE.
CC -1- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -1- SUBUNIT: TETRAMER CONSISTING OF TWO ALPHA (CATALYTIC) AND TWO BETA
CC (FUNCTION NOT KNOWN) CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08471; AAA22384.1; -
DR EMBL; Z99112; CAB13549.1; -
DR PIR; C46665; C46665.

DR Subtilist; BG10784; dapG.
DR InterPro; IPR001048; As_kinase.
DR InterPro; IPR00260; Asp_kin_monofn.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00696; aakinas; 1.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Transferase; Kinase; Diaminopimelate biosynthesis;
KW Lysine biosynthesis; Complete proteome.
FT CHAIN 1 404 ASPARTOKINASE I ALPHA SUBUNIT.
FT CHAIN 245 404 ASPARTOKINASE I BETA SUBUNIT
FT CHAIN 399 399 E -> V (IN REF. 1).
FT CONFLICT 399 399 (BY SIMILARITY).
SQ SEQUENCE 404 AA; 42978 MW; 49A6DA70D70047C4 CRC64;
Query Match 24.3%; Score 532; DB 1; Length 404;
Best Local Similarity 36.5%; Pred. No. 3e-30;
Matches 145; Conservative 62; Mismatches 122; Indels 68; Gaps 11;
QY 1 VQYGGSSVADAERIRRAERIVATKKQNDVVVVSAMGD-----TTDDLDDLAQQVCP 55
DB 5 VQKGGTSVQKDKGRKALGHIKEAISEGYKVVVVSNMGRKGPYATDSLLGLYGQS 64
QY 56 APPRELDMLTAGERISNALVAIBSLGAHARSFTGSQAGVITTGTHGNAKIIDVTPG 115
DB 65 AISPREQDILLSCGETISSVFTSMLLDNGVKAALITGAQAQGLTNDQHTNAKIEMKPE 124
QY 116 RLQTALEGRVYVAGFQGVQSDIKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDDGI 175
DB 125 RLFSVLANDHADVAVAGFQCAT-EGKDTTIGRGSDTSAAALGAADVADYDITFDVEGV 183
QY 176 PSADPRIVRNARKLDTVFEEMLMAACAKVLMLRCVEYARRHNPVHVRSYSDRPGT 235
DB 184 MTADPRVVENAKPLPVVYITICNLAYQCAKVISFRAVEIAQAKVPVRVSTYSNDKGT 243
QY 236 VV-----VGSIKDVPNEDILITGAH--DRSEAKVTI-VGLDIPGYAAKVFEEAVRR 285
DB 244 LVTSHSSKVGVS---DVFERLITGIAHKVDVTFKVPKIQGVNV---QTEVFKAFA-- 294
QY 286 RSQHRHGAERLQGRGRQRDRHLLPHTGSPPPMKWTRSE-----TRSASTQLL 336
DB 295 -----NAGISVDFFNITPSEIVVTVAGNKTETAQRILM 327
QY 337 ---YDHI-----GKVSLLIGAGMRSHPGVTFATFCEALA 366
DB 328 DMGYDPMVTRNCARKVSAVGAGIMGVPGVTSKIVSALS 364
RESULT 15
AK_METUA STANDARD; PRT; 473 AA.
ID AK METUA
AC Q57991;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN M00571.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
CX NCBI_TaxID=2190;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.E., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

Search completed: November 21, 2003, 16:04:16
Job time : 5.99955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 20.8635 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-4
Perfect score: 2193
Sequence: 1 VQYGGSSVADARRVAE.....SAATRRPCTRGDGRWACQ 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258032604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1545.5	70.5	421	16	Q9CB77	Q9cb77 mycobacteri
2	1272.5	58.0	421	2	Q8KQ29	Q8kq29 streptomyc
3	1267.5	57.8	421	2	Q9RQ25	Q9rq25 amycolatops
4	1232.5	56.2	421	2	Q93C57	Q93c57 corynebacte
5	1228.5	56.0	421	2	Q9AE85	Q9ae85 nocardia la
6	1224.5	55.8	421	2	Q93C55	Q93c55 corynebacte
7	1141	52.0	425	16	Q9XA17	Q9xa17 streptomyc
8	1127	51.4	423	2	Q8GIQ5	Q8giq5 streptomyc
9	821	37.4	434	2	Q93TX4	Q93tx4 stigmatella
10	816.5	37.2	600	16	P74569	P74569 synchocyst
11	803	36.6	405	15	Q8F865	Q8f865 leptospira
12	798.5	36.4	408	2	Q8RU02	Q8ru02 methylobaci
13	786.5	35.9	599	16	Q8DHW0	Q8dhw0 synchococc
14	776.5	35.4	606	16	Q8YR14	Q8yr14 anabaena sp
15	773.5	35.3	254	16	Q8G6Y6	Q8g6y6 bifidobacte
16	761	34.7	405	16	Q9JTN3	Q9jtn3 neisseria m

17	760	34.7	405	15	Q9JYN6	Q9jyn6 neisseria m
18	757.5	34.5	412	16	Q9K8B0	Q9k8b0 bacillus ha
19	742.5	33.9	400	16	Q9PHI4	Q9ph4 campylobact
20	724.5	33.0	401	16	Q9XIK6	Q9xik6 thermotoga
21	724.5	33.0	739	16	Q9WZ17	Q9wz17 thermotoga
22	707.5	32.3	416	16	Q8Y073	Q8y073 ralstonia s
23	707	32.2	404	16	Q92CH4	Q92ch4 listeria in
24	705	32.1	415	16	Q9A9W8	Q9a9w8 caulobacter
25	703	32.1	404	16	Q8Y7N9	Q8y7n9 listeria mo
26	694.5	31.7	423	16	Q8Y097	Q8y097 brucella me
27	690.5	31.5	423	16	Q8FYK1	Q8fyk1 brucella su
28	664.5	30.3	418	2	Q9EV19	Q9ev19 bradyrhizob
29	663	30.2	417	16	Q98G91	Q98g91 rhizobium l
30	656	29.9	431	16	Q8URC0	Q8urc0 agrobacteri
31	652.5	29.8	424	16	Q92MK3	Q92mk3 rhizobium m
32	632	28.8	395	16	Q9KH4	Q9kuh4 vibrio chol
33	605	27.6	418	16	Q8EBS2	Q8ebs2 shewanella
34	602	27.5	398	16	Q8RA48	Q8ra48 thermococ
35	585	26.7	370	16	Q8DC48	Q8dc48 vibrio vuln
36	585	26.7	401	16	Q99U91	Q99u91 staphylococ
37	584	26.6	401	2	Q9EV36	Q9ev36 staphylococ
38	584	26.6	401	16	Q8NW57	Q8nws7 staphylococ
39	575	26.2	410	16	Q8CSN8	Q8csn8 bacillus ha
40	562	25.6	411	16	Q9RA90	Q9ra90 oceanobacil
41	536	24.4	411	16	Q8EQT0	Q8eqt0 oceanobacil
42	524	23.9	403	16	Q92BR9	Q92br9 listeria in
43	520	23.7	401	16	Q92GF9	Q92gf9 rickettsia
44	506	23.1	403	16	Q8Y765	Q8y765 listeria mo
45	467	21.3	398	16	Q8XJ56	Q8xj56 clostridium

ALIGNMENTS

RESULT 1
Q9CB77 PRELIMINARY; PRT; 421 AA.
ID Q9CB77; AC Q9CB77; DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Aspartokinase.
GN ASK OR ML2323.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IN; RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
DR ENBL; AL583925; CAC31839.1; --
DR Leproma; ML2323; --
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR Pfam; PF00659; aakinas; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfams; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 421 AA; 44663 MW; 4A3BA44D1CEEFE06 CRC64;

```

Query Match      70.5%; Score 1545.5; DB 16; Length 421;
Best Local Similarity 80.4%; Pred. No. 3e-96;
Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 1 VOKYGGSSVADAEIRIRVAERIVATKKQGNVVVVVSAMGDTTDDLLDLAQOVCPPPPR 60
DB 5 VOKYGGSSVADAEIRIRVAERIVATKKQGNVVVVVSAMGDTTDDLLDLAQOVCPPPPR 64
QY 61 ELDMLITAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
DB 65 ELDMLITAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 124
QY 121 LEBGRVVLVAGQGVSDTKVTTILGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 180
DB 125 LDEGRVVLVAGQGVSDTKVTTILGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 184
QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 240
DB 185 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 244
QY 241 IKDVPMEDILTVGAHDSERAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE----- 295
DB 245 IKDVPMEDILTVGAHDSERAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE----- 295
QY 296 RLQGRGRDRHLLH---LPQSGPPPKWNTSRSETRASTOLLYDDHIGKVSILGAGR 352
DB 296 VLQNVSKVEDGKTDTFTCSRDGSLIAVAKGSLRDEIGFTQLLYDDHIGKVSILGAGR 355
QY 353 SHPGVTATFCEALAAVGNIELISTSEDRGCCAAT 389
DB 356 SHPGVTATFCEALAAVGNIELISTSEIRISVLCDRT 392

RESULT 2
Q8KQ29 ID Q8KQ29 PRELIMINARY; PRT; 421 AA.
AC Q8KQ29;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Aspartate kinase alpha subunit (EC 2.7.2.4) (Aspartokinase).
GN ASKA.
OS Streptomyces clavuligerus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1901;
RN [1]
RP SEQUENCE FROM N.A.
RA Tunica S.;
RT "Isolation and characterization of the aspartokinase and aspartate semialdehyde dehydrogenase genes from cephamycin C-producer Streptomyces clavuligerus."
RT Thesis (2002), University of Middle East Technical University, Ankara, Turkey.
RL
CC -|- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
CC -|- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL; AY112728; AAM54736.1; -.
CC InterPro; IPR001048; Aa_kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR001341; Aspartate_kinase.
CC InterPro; IPR005260; Asp_kin_monofn.
CC InterPro; IPR001057; Glu_kinase.
CC Pfam; PF00696; aakinas; 1.
CC Pfam; PF01842; ACT; 2.
CC PRINTS; PR00474; GLUSKINASE.
CC TIGRFS; TIGR00657; asp_kinases; 1.
CC TIGRFS; TIGR00656; asp_kin_monofn; 1.
CC PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44399 MW; 08262D81045735C1 CRC64;

Query Match      58.0%; Score 1272.5; DB 2; Length 421;
Best Local Similarity 66.1%; Pred. No. 7.5e-78;
Matches 257; Conservative 51; Mismatches 60; Indels 21; Gaps 4;

QY 1 VOKYGGSSVADAEIRIRVAERIVATKKQGNVVVVVSAMGDTTDDLLDLAQOVCPPPPR 60
DB 5 VOKYGGSSLESADRIKRAERIVATKKAGNEVVVVVCSAMGDTTDELLDLAQOVPVPPR 64
QY 61 ELDMLITAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGHNAKIIDVTPGRLOTA 120
DB 65 EMDMLITAGERISNALVAMAIESAQAQANSFTGSQAGVITTAHGNARIIDVTPSEVSA 124
QY 121 LEBGRVVLVAGQGVSDTKVTTILGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 180
DB 125 LEOGYVALVAGQGVQAQDNDITTLGRGSDTTAVAMAAALNADVCEIYSDVDGYSDAP 184
QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVGS 240
DB 185 RIVPNARKLDSVPEEMLEMAAGSKILHRSVEYARRYGVFIRVSSYSDRPGTVVGS 244
QY 241 IKDVPMEDILTVGAHDSERAKVTIVGPDIPGYAAKVFRAVARRRRQHRHGAEE----- 295
DB 245 IEEIPVQALITGVADHDSERAKVTIVGPDHAGAAARIPVIA-----DAEIDIM 295
QY 296 RLQ-----GRGRDRHLLHLLPQSGPPPKWNTSRSETRASTOLLYDDHIGKVSILGAG 350
DB 296 VLQNVNTSGRTD---ITTLKANGPKAVASLEKIKELGFSVLYDDHVGKVSILGAG 353
QY 351 MRSHPGVTATFCEALAAVGNIELISTSE 379
DB 354 MRSHPGVTATFCEALSKGVNIEINTSE 382

RESULT 3
Q9RQ25 ID Q9RQ25 PRELIMINARY; PRT; 421 AA.
AC Q9RQ25;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Aspartokinase subunit A (EC 2.7.2.4) (Aspartate kinase).
GN ASKA.
OS Amycolatopsis mediterranei (Nocardia mediterranei).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
OX NCBI_TaxID=33910;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=U32;
RX MEDLINE=9453302; PubMed=10521665;
RX Zhang W., Jiang W., Zhao G., Yang Y., Chiao J.;
RT "Sequence analysis and expression of the aspartokinase and aspartate semialdehyde dehydrogenase operon from rifamycin SV-producing amycolatopsis mediterranei."
RT Gene 237:413-419(1999).
RL
CC -|- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
CC -|- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL; AF134837; RAD49567.1; -.
CC InterPro; IPR001048; Aa_kinase.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR001341; Aspartate_kinase.
CC InterPro; IPR005260; Asp_kin_monofn.
CC InterPro; IPR001057; Glu_kinase.
CC Pfam; PF00696; aakinas; 1.
CC Pfam; PF01842; ACT; 2.
CC PRINTS; PR00474; GLUSKINASE.
CC TIGRFS; TIGR00657; asp_kinases; 1.
CC TIGRFS; TIGR00656; asp_kin_monofn; 1.
CC PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44393 MW; 633D9C2D023145E9 CRC64;

```

```
Query Match          57.8%; Score 1267.5; DB 2; Length 421;
Best Local Similarity 65.3%; Pred. No. 1.6e-77;
Matches 254; Conservative 53; Mismatches 61; Indels 21; Gaps 4;

QY 1 VKYGGSSVADAERIRVAERIVATKKQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 185 RVVPDAKLDVTVEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 244

QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 290
D 245 IEIPIEQALITGVAHDSRSEAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 290

QY 296 RLQG-----RGRDRHLLHLLPQTSPPPPKNNWTRSETRSASTOLLYDDHIGKVSILIGAG 350
D 296 VLQNVSTVSGRTD--ITFTLSKANGAKAVKELEKVAEIGFESVLYDDHVGKVSVVVGAG 353

QY 351 MRSHPGVGTATFCEALAAVGNIELISTSE 379
D 354 MRSHPGVGTATFCEALAAVGNIELISTSE 382

RESULT 4
Q93C57 PRELIMINARY; PRT; 421 AA.
AC Q93C57
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Feedback-resistant aspartokinase LysC alpha subunit (EC 2.7.2.4)
DE (Aspartate kinase)
OS Corynebacterium crenatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=168810;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD945;
RA Liu Y., Ding J., Wang J.;
RT "Cloning and sequence analysis of aspartokinase genes from
RT Corynebacterium crenatum."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL; AF14084; AAL07807.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44738 MW; 473A19409C0215B6 CRC64;

Query Match          56.2%; Score 1232.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 3.7e-75;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 1 VKYGGSSVADAERIRVAERIVATKKQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 185 RVVPDAKLDVTVEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 244

QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 290
D 245 IEIPIEQALITGVAHDSRSEAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 290

QY 296 RLQG-----RGRDRHLLHLLPQTSPPPPKNNWTRSETRSASTOLLYDDHIGKVSILIGAG 350
D 296 VLQNVSTVSGRTD--ITFTLSKANGAKAVKELEKVAEIGFESVLYDDHVGKVSVVVGAG 353

QY 351 MRSHPGVGTATFCEALAAVGNIELISTSE 379
D 354 MRSHPGVGTATFCEALAAVGNIELISTSE 382

RESULT 5
Q9AE85 PRELIMINARY; PRT; 421 AA.
AC Q9AE85;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aspartokinase alpha subunit (EC 2.7.2.4) (Aspartate kinase).
DE ASK.
GN ASK.
OS Nocardia lactamdurans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.
OX NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hernandez-Rico V.;
RL Thesis (2000), Department of Area of Microbiology, University of Leon, Leon, Spain.
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL; AJ298904; CAC37034.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate kinase.
DR InterPro; IPR001057; Glu_Skinase.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRFAMs; TIGR00657; asp_kinases; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44141 MW; 802FBE02D01B7F95 CRC64;

Query Match          56.0%; Score 1228.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 7e-75;
Matches 250; Conservative 52; Mismatches 66; Indels 21; Gaps 4;

QY 1 VKYGGSSVADAERIRVAERIVATKKQNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 120
D 5 VKYGGSSLESADRIKRVAEIRVATKKAGNDVVVVVCSAMGDTTDELLELAAVNPVPPAR 64

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 124

QY 121 LEEGRVVLVAGFGVGSQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 180
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 125 LDQGYIALVAGFGVGAQDITKDVTTILGRGSGDTTAVAMAAALGADVCEIYTDVGDGIFSADP 184

QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 240
D 185 RVVPDAKLDVTVEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVGVS 244

QY 241 IKDVPMDPILTGVAHDSRSEAKVTIVGLPDPGVAAKVFRAVA-----RRRQHR 290
D 305 DGTDTITCTPRSDGRRAMEILKKPQVG-----NW-----TNVLYDDQVGVSLV 350

QY 348 GAGMESHGPGVGTATFCEALAAVGNIELISTSE 379
D 351 GAGMESHGPGVGTATFCEALAAVGNIELISTSE 382
```

Db 65 EMDMLTAGDGSINALVAVAIQAQAQWSFTGSGAGVVTAVHGNARIIDVTPSRVSR 124
 QY 121 LBEGRVLVAGFGQVSQDTKVTTILGRGSGDITAVAMAAAGADYCEIYTDVDFGSADP 180
 Db 125 LQGVVALVAGFGQVADTKDITILGRGSGDITAVAMAAAGADYCEIYSDVDGYSADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSDRPGTVVGS 240
 Db 185 RIVPARKLDSVPYEEMLEMAAGSKILHLRSVESARRYGVPIRVSSYSKPGKTTVGS 244
 QY 241 IKDVPMDPILITGVADRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAAB----- 295
 Db 245 IBEIPVEQALITGVADRSEAKVTIVGPDHAGAARIFRVIA-----DAEIDIDM 295
 QY 296 RLQ-----GRGQRDRHLHLPLQTPGPPPKWNTSRSEASTOLLYDDHIGKVSILGAG 350
 Db 296 VLQNVSTSSGRD--ITFLSKANGPKRAVASLEKIKELGFSVLYDDHVGKVSILGAG 353
 QY 351 MRSHPGVATFCEALAAVGNIELISTSE 379
 Db 354 MRSHPGVATFCEALCKVGNIEINTSE 382

RESULT 6

Q93C55 ID Q93C55 PRELIMINARY; PRT; 421 AA.
 AC Q93C55;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Aspartokinase Lysc alpha subunit (EC 2.7.2.4) (Aspartate kinase).
 OS Corynebacterium crenatum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=168810;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1542;
 RA Liu Y., Ding J., Wang J.;
 RT "Cloning and sequence analysis of aspartokinase genes from Corynebacterium crenatum."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-ASPARTATE.
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 DR EMBL; AF414085; AAL07809.1; -;
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR001341; Aspartate kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR Pfam; PF00696; aakinasase; 1.
 DR Pfam; PF01842; ACT; 2.
 DR TIGRfam; TIGR00657; asp_kinases; 1.
 DR TIGRfam; TIGR00656; asp_kin_monofn; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Kinase; Transferase.
 SQ SEQUENCE 421 AA; 44754 MW; E36B4D008IDE0827 CRC64;

Query Match 55.8%; Score 1224.5; DB 2; Length 421;
 Best Local Similarity 64.0%; Pred. No. 1.3e-74;
 Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
 QY 1 VQYGGSSVADAERIRRVARRIVATKQGNDDVVVVSAMGDTTDDLLDLAQVCPAPP 60
 Db 5 VQYGGSSLESARIRNVAERIVATKKGAGNDVVVCSAMGDTTDELELAANVPVPPAR 64
 QY 61 ELDMLLTAGERISNALVAVAIQAQAQWSFTGSGAGVVTAVHGNARIIDVTPSRVSR 120
 Db 65 EMDMLLTAGERISNALVAVAIQAQAQWSFTGSGAGVVTAVHGNARIIDVTPSRVSR 124
 QY 121 LBEGRVLVAGFGQVSQDTKVTTILGRGSGDITAVAMAAAGADYCEIYTDVDFGSADP 180

Db 125 LBEGRVLVAGFGQVSQDTKVTTILGRGSGDITAVAMAAAGADYCEIYSDVDGYSADP 184
 QY 181 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSYSDRPGTVVGS 240
 Db 185 RIVPARKLDSVPYEEMLEMAAGSKILHLRSVESARRYGVPIRVSSYSKPGKTTVGS 244
 QY 241 IKDVPMDPILITGVADRSEAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAAB----- 295
 Db 245 IBEIPVEQALITGVADRSEAKVTIVGPDHAGAARIFRVIA-----DAEIDIDM 295
 QY 296 RLQ-----GRGQRDRHLHLPLQTPGPPPKWNTSRSEASTOLLYDDHIGKVSILGAG 350
 Db 296 VLQNVSTSSGRD--ITFLSKANGPKRAVASLEKIKELGFSVLYDDHVGKVSILGAG 353
 QY 351 MRSHPGVATFCEALAAVGNIELISTSE 379
 Db 354 MRSHPGVATFCEALCKVGNIEINTSE 382

RESULT 7

Q9XA17 ID Q9XA17 PRELIMINARY; PRT; 425 AA.
 AC Q9XA17;
 DT 01-NOV-1999 (TREMELrel. 12, Created)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Putative aspartokinase.
 GN ASK OR SCO3615 OR SC66T3.26C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=843436;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RC MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL; AL539117; CAB45482.1; -;
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR001341; Aspartate kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR Pfam; PF00696; aakinasase; 1.


```
RESULT 12
Q8RU02 PRELIMINARY; PRT; 408 AA.
AC Q8RU02;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Aspartate kinase (EC 2.7.2.4) (Aspartokinase).
GN LYSC.
OS Methylobacillus glycozenes.
OC Bacteria; Proteobacteria; Betaproteobacteria; Methylobacterales;
OC Methylobacillaceae; Methylobacillus.
CX NCBI_TaxID=406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 21371;
RA Motocoma H.; Anazawa H.;
RT "Analysis of the nucleotide sequence of lysC of Methylobacillus
glycozenes.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
ASPARTATE.
CC -|- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
DR EMBL; AB038354; BAB8486.1; -.
DR InterPro; IPR001048; Aa_kinase.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001341; Aspartate_kinase.
DR InterPro; IPR005260; Asp_kin_monofn.
DR Pfam; PF00696; aak_kinase; 1.
DR Pfam; PF01842; ACT; 2.
DR TIGRfams; TIGR00657; asp_kinases; 1.
DR TIGRfams; TIGR00656; asp_kin_monofn; 1.
DR PROSITE; PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 408 AA; 44328 MW; 3894P2315AGEBD75 CRC64;

Query Match 36.4%; Score 798.5; DB 2; Length 408;
Best Local Similarity 41.7%; Pred. No. 6.5e-46;
Matches 168; Conservative 78; Mismatches 114; Indels 43; Gaps 3;

QY 1 VKYGGSSVADAERIRVAERIVATKKGNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VKYGGTSVANPERIGNVARRVAEHKALGOVVVVVSAMSGEINRLIGLAKVWPNPFR 64
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSTGQAGVITGTHGNAKIIDVTPGRLQTA 120
DB 65 ELDVLVSTGQVTIALLSMALMELGVKAKSYTSQVSIVTDNAFTKARILKIDEEIRKD 124
QY 121 LEEGRVVLVAGFGQVSDTK-DVTTTLGRGGSDTTAVAMAALGADVCEIYTDVGI SADP 180
DB 125 LDDGVVVVVTGFGQVDEH-GNITTLGRGGSDTTGVALATALKADEQIYTDVGVITTD 183
QY 181 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDREGTVVVG 240
DB 184 RVWPEARRLKSVTPEEMLEMASLGSKVLRVSVEFAGKYKVKLRVLSSPEEGEGLITF 243
QY 241 IKQVPMEDPILTVGAHDSERAKTIVGLPIPGYAAKVFAVARRR----- 286
DB 244 EDDNNMEQPIISGIAFNDEAKITVLGPDPGPGIAYQILGFDVADINVDIIONVAADG 303
QY 287 -----RQHRGAERLQGRGQDRHLLHLLPOTSGPPPKWTRSETRSASTQLLYDDH 340
DB 304 TTDFTFTVHKNDLQXALDILRDVKVQGH-----GAREIIGDDK 341
QY 341 IGKVLGAGNRSHPGVTATFCEALAAVGNIELISTSEDSQRS 383
DB 342 TAKVSVVGVGNRSHVGIQVSMFRTLABEGINQIMISTSEIQIS 384

RESULT 13
Q8DHWO PRELIMINARY; PRT; 599 AA.
AC Q8DHWO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Aspartate kinase.
GN TLR1833.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
CX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Nakamura Y., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130 (2002).
DR EMBL; AF005375; BAC09385.1; -.
KW Kinase; Complete proteome.
SQ SEQUENCE 599 AA; 62817 MW; 93D7D3ACDE2802A1 CRC64;

Query Match 35.9%; Score 786.5; DB 16; Length 599;
Best Local Similarity 45.7%; Pred. No. 7e-45;
Matches 184; Conservative 67; Mismatches 129; Indels 23; Gaps 7;

QY 1 VKYGGSSVADAERIRVAERIVATKKGNDVVVVVVSAMGDTTDDLLDLAQVCPAPPPR 60
DB 5 VKYGGTSVGSVERIQAVRVKATVAGHQVVVVVSAMGKTTDSLVLQLAISAIDRPSQR 64
QY 61 ELDMLLTAGERISNALVAMAEISGAHARSTGQAGVITGTHGNAKIIDVTPGRLQTA 120
DB 65 EMDMLLSTGEQVSIALLTMALHALGEPAISLTGAQGVITPEPAHTRARILHIETORLBRH 124
QY 121 LEEGRVVLVAGFGQVSDTK-DVTTTLGRGGSDTTAVAMAALGADVCEIYTDVGI SAD 179
DB 125 LKAGGVVVVAGFGQITAAATDFEVTTLGRGGSDTSAVALAALRADCCETIYTDVPGILTA 184
QY 180 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDREGTVVVG 239
DB 185 PELVFNQAQLMSEITCDEMLELASLGAKVLPRAVEIARNYGVDLVVRSSWTDGPTRVIA 244
QY 240 -----SIKQVPMEDPILTVGAHDSERAKTIVGLPIPGYAAKVFAVARRR----- 287
DB 245 PARPPRPVENVLELQKPV-DGVALDQAKVALLRVADRPGVAAQLFGLARQNLVDVLI 303
QY 288 QHRHGAERLQGRGQDRHLLH--LLPOTSGPPPKWTRSETRSASTQLLYDDHIGKVS 345
DB 304 QSIH-----EGQTNDAFTVQKSVLKQAEALV-AFYPLSPRVEETDVLVDADI AKVS 356
QY 346 LIGAGNRSHPGVTATFCEALAAVGNIELISTSEDSQRS 388
DB 357 ITGAGMIGRPGVAAQMFSAALAAAGINLQIMISTSEVNVSVCTVAA 399

RESULT 14
Q8YR14 PRELIMINARY; PRT; 606 AA.
AC Q8YR14;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Aspartate kinase.
GN ALR3644.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
```

RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120.";
 RL DNA Res. 8:205-213 (2001).
 DR EMBL; AP003593; BAB75343.1; --
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR002912; ACT
 DR InterPro; IPR001341; Aspartate_kinase.
 DR InterPro; IPR005260; Asp_kin_monofn.
 DR Pfam; PF01842; ACT; 4.
 DR TIGRFAMs; TIGR00657; asp_kinases; 1.
 DR TIGRFAMs; TIGR00656; asp_kin_monofn; 1.
 DR PROSITE; PS00324; ASPARTOKINASE; 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 606 AA; 54382 MW; 932414217271D121 CRC64;

Query Match 35.4%; Score 776.5; DB 16; Length 606;
 Best Local Similarity 43.2%; Pred. No. 3.3e-44;
 Matches 175; Conservative 74; Mismatches 115; Indels 41; Gaps 7;
 QY 1 VQYGGSSVADAEIRRVRAERIVATKQGNVWVVSAMGDTTDDLLDLAQOVCPPPPR 60
 DB 5 VQKFGSSVGSVERIQAVQRYKTVKAGNSLVVVSAMGKTTDGLVKLANEISRNPNR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIVTGPRLQTA 120
 DB 65 EMDMLLTAGEQVITALLSWALQELGQPAISMTGQVGLVTEAETHRIILHIERVEGH 124
 QY 121 LEEGRVVLVAGFQVSQD-TKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGFSAD 179
 DB 125 IKEGRVWVWVAGFQGSISLEITTLGRGSDTSVALAALQADFCEIYTDVDFGILTDD 184
 QY 180 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVV- 238
 DB 185 RIVFEAQLMGEITCNEMLEIASLGAKVLHPRAVEIARNYGNPLVVKSSWTDDEGTWTT 244
 QY 239 -----GSIKDVPMEDPILTVGAHRSEAKVTIVGLPDIPGVAAKVFRVARR-----R 287
 DB 245 KPQGRSLINLELARPV-DVVELDTQAKVALLRVPKPGVAAKLFGELISQKQVDVLI 303
 QY 288 QHRHGAERLQGRG-----RQDRHLHLHPQTSGPPPPKWNWTFSETRSASTQ 334
 DB 304 QSIH-----EGNSNDIAFTVTTPIPKRAEAAVAIAAPSLRSP-----SHPKSDAE 349
 QY 335 LLYDDHIGKSLIGAGMRSHPGVTATFCEALAAAGVNIETISTSE 379
 DB 350 VVVEQNIKAVSIAGAGMIGRPGVAQMFPATLAEAGVNIQIMISTSE 394

RESULT 15
 Q8G6Y6 PRELIMINARY; PRT; 254 AA.
 AC Q8G6Y6;
 DT 01-MAR-2003 (TRENDELrel. 23, Created)
 DT 01-MAR-2003 (TRENDELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENDELrel. 23, Last annotation update)
 DE Aspartokinase.
 GN ASKA OR BL0494.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN [1]_TaxID=216816;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arigoni F.;

RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation
 to the human gastrointestinal tract.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
 DR EMBL; AB014669; AAN24323.1; --
 KW Kinase; Complete proteome.
 SQ SEQUENCE 254 AA; 27041 MW; 28D9F48C68822225 CRC64;
 Query Match 35.3%; Score 773.5; DB 16; Length 254;
 Best Local Similarity 64.3%; Pred. No. 1.7e-44;
 Matches 153; Conservative 38; Mismatches 46; Indels 1; Gaps 1;
 QY 1 VQYGGSSVADAEIRRVRAERIVATKQGNVWVVSAMGDTTDDLLDLAQOVCPPPPR 60
 DB 5 VQKFGSSVADTESIKRVAKRVVETEKGNKVAVVVSAMGDTTDDLLDLQALSIDSNPPR 64
 QY 61 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIVTGPRLQTA 120
 DB 65 EMDMLLTAGERISMSLLAMAIHAEGSRAHSFTGQAGFTTDARYGAARIKAVRPDRVXNA 124
 QY 121 LEEGRVVLVAGFQVSQD-TKDVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFGFSADP 180
 DB 125 LSLGDIIVAGFQGINA-KGDATTLGRGSDTSVALAVALGADICEIYTDVDFGIFTADP 183
 QY 181 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSDRPGTVV 238
 DB 184 RIVFSARRIPSIDYESILEMASCCKVLAALRCVEYAOQRFNMPHVRSSFSRPPGLTV 241

Search completed: November 21, 2003, 16:08:08
 Job time : 22.8635 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 : Search time 1.39184 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence:

1 VIDLIIGTSPSWEQAAAEAV.....KITRIKLEVSFQWRPAQRR 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	324	100.0	66	3	US-08-818-112-77
2	324	100.0	66	4	US-08-818-111-78
3	324	100.0	66	4	US-09-056-556-77
4	324	100.0	66	4	US-09-072-596-78
5	324	100.0	802	4	US-09-056-556-214
6	324	100.0	802	4	US-09-072-596-209
7	324	100.0	802	4	US-09-072-596-346
8	69	21.3	92	4	US-09-252-991A-21917
9	61.5	19.0	657	4	US-09-300-909-16
10	58	17.9	446	4	US-09-252-991A-25318
11	57.5	17.7	500	2	US-08-031-392-7
12	57.5	17.7	500	3	US-09-299-549-7
13	57.5	17.7	500	4	US-09-610-417-7
14	57.5	17.7	633	4	US-09-252-991A-26229
15	57.5	17.7	900	4	US-09-328-352-8072
16	56	17.3	286	4	US-09-071-035-176
17	56	17.3	305	4	US-09-071-035-174
18	56	17.3	340	4	US-08-252-991A-24373
19	56	17.3	541	4	US-09-252-991A-27169
20	56	17.3	2548	4	US-09-172-422-1
21	55.5	17.1	613	4	US-09-149-727-5
22	55.5	17.1	651	4	US-09-715-858-2
23	55	17.0	597	4	US-09-252-991A-25218
24	55	17.0	1705	4	US-08-669-785-4
25	54.5	16.8	286	4	US-09-328-352-7001
26	54.5	16.8	648	4	US-09-715-858-4
27	54	16.7	360	4	US-09-252-991A-16702

28	54	16.7	450	4	US-09-252-991A-29611	Sequence 29611, A
29	53.5	16.5	363	3	US-09-046-086-2	Sequence 2, Appli
30	53.5	16.5	363	4	US-09-524-643-2	Sequence 2, Appli
31	53.5	16.5	384	4	US-09-252-991A-27463	Sequence 27463, A
32	53.5	16.5	430	4	US-09-252-991A-31388	Sequence 31388, A
33	53.5	16.5	462	1	US-08-336-408B-2	Sequence 2, Appli
34	53.5	16.5	462	3	US-08-764-870-6	Sequence 2, Appli
35	53.5	16.5	462	3	US-08-980-115-6	Sequence 6, Appli
36	53.5	16.5	462	5	PCT-US91-00399-2	Sequence 6, Appli
37	53.5	16.5	626	4	US-09-252-991A-20601	Sequence 2, Appli
38	53	16.4	733	4	US-09-252-991A-19533	Sequence 20601, A
39	53	16.4	733	3	US-09-193-983-6	Sequence 19533, A
40	53	16.4	871	4	US-09-134-001C-3979	Sequence 3979, Ap
41	53	16.4	985	4	US-09-252-991A-30698	Sequence 30698, A
42	52.5	16.2	551	2	US-09-067-351-2	Sequence 2, Appli
43	52.5	16.2	551	3	US-09-360-490-2	Sequence 2, Appli
44	52.5	16.2	745	4	US-09-107-532A-5687	Sequence 5687, Ap
45	52.5	16.2	824	4	US-09-252-991A-32329	Sequence 32329, A

ALIGNMENTS

RESULT 1

US-08-818-112-77
; Sequence 77, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-77

Query Match 100.0%; Score 324; DB 3; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 2
US-08-818-111-78
; Sequence 78, Application US/0881811
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-818-111-78
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66
RESULT 3
US-09-056-556-77
; Sequence 77, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

QY 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 3
US-09-056-556-77
; Sequence 77, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

QY 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66
RESULT 4
US-09-072-596-78
; Sequence 78, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-056-556-77
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSPTSEWQAAAEVQARQSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

US-09-072-596-78
; Sequence 78, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-072-596-78
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 5
US-09-056-556-214
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-072-596-78
Query Match 100.0%; Score 324; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 6
US-09-072-596-209
; Sequence 209, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vadvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-072-596-209
Query Match 100.0%; Score 324; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 5.9e-36;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

TOPLOGY: linear
Query Match 100.0%; Score 324; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 5.9e-36;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74
```

Db 69 RPAQR 74

RESULT 7
US-09-072-596-346
; Sequence 346, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-072-596-346

Query Match 100.0%; Score 324; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 5.9e-36;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
|||||
Db 9 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 68
|||||

QY 61 RPAQR 66
|||||
Db 69 RPAQR 74

RESULT 8
US-09-252-991A-21917
; Sequence 21917, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21917
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21917

Query Match 21.3%; Score 69; DB 4; Length 92;
Best Local Similarity 25.4%; Pred. No. 0.038;
Matches 15; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVSFKM 60
|||||
Db 31 IELVGSKTSIEDAINNALAAEAQSIQHLWEFEVVDTRGHIENGAVGHVQVTLKVGFR 89
|||||

RESULT 9
US-09-300-909-16
; Sequence 16, Application US/09300909
; Patent No. 6306580
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PREPARATION OF HUMAN PAPILLOMAVIRUS E1 HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND METHOD THEREFOR
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,909
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/083,942
; FILING DATE: 01-MAY-1998
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-300-909-16

Query Match 19.0%; Score 61.5; DB 4; Length 657;
Best Local Similarity 29.2%; Pred. No. 5.3;
Matches 19; Conservative 14; Mismatches 29; Indels 3; Gaps 1;

QY 1 VIDIGTSPTEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITTYRIKLEVS 57
|||||
Db 47 MVDFITQTFCEQAELETAQALFHAQVEVNDQAQVHLVKRKFAGGSGSTENSLGERLEVD 106
|||||

QY 58 FQVRP 62
|||||
Db 107 TELSP 111

RESULT 10
US-09-252-991A-25318
; Sequence 25318, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-610-417-7

Query Match 17.7%; Score 57.5; DB 4; Length 500;
Best Local Similarity 43.1%; Pred. No. 13;
Matches 22; Conservative 9; Mismatches 13; Indels 7; Gaps 4;

Qy 13 EQAAAEAVQEAR--DSVDIRVARVIEQDMVDSAGKITRYIKLEVSFKMR 61
Db 227 EAAKALQTLRGWSVDR-EVAEIRQEDAEKAAAGFISV-LKL---FMR 272

US-09-328-352-8072
Sequence 8072, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8072
LENGTH: 900
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8072

Query Match 17.7%; Score 57.5; DB 4; Length 900;
Best Local Similarity 31.7%; Pred. No. 28;
Matches 19; Conservative 9; Mismatches 21; Indels 11; Gaps 2;

Qy 13 EQAAAEAV-----QRADSVDIRVARVIEQDMVDSAGKITRYIKLEVSFKMR 63
Db 122 EQQAREAAEQKAEQLQTEQKAKATLDAAAH--QDSAAQSAKAAVVVKRGGGTVRPA 179

Search completed: November 21, 2003, 16:11:38
Job time : 2.39184 secs

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-610-417-7

Query Match 17.7%; Score 57.5; DB 4; Length 500;
Best Local Similarity 43.1%; Pred. No. 13;
Matches 22; Conservative 9; Mismatches 13; Indels 7; Gaps 4;

Qy 13 EQAAAEAVQEAR--DSVDIRVARVIEQDMVDSAGKITRYIKLEVSFKMR 61
Db 227 EAAKALQTLRGWSVDR-EVAEIRQEDAEKAAAGFISV-LKL---FMR 272

RESULT 14
US-09-252-991A-26229
Sequence 26229, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107195.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26229
LENGTH: 833
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26229

Query Match 17.7%; Score 57.5; DB 4; Length 633;
Best Local Similarity 31.7%; Pred. No. 18;
Matches 20; Conservative 8; Mismatches 20; Indels 15; Gaps 3;

Qy 1 VIDICTSTPTSEQAAAEAVQ-----RADSVDIRVARVIEQ-DM-AVDSA 45
Db 257 VADLVGTHPDQVEAAIRVVRVENIDQALLAVHARPDLPDRIAHAAQHLDMGAVDIA 316

Qy 46 GKI 48
Db 317 GAV 319

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 3.99592 seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence: 1 VIDIIQTSPTEWQAAAEAV.....KITRIKLEVSFKMRPAQPR 66

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	324	100.0	66	18	AAW32433
2	324	100.0	66	18	AAW32365
3	324	100.0	66	19	AAW81668
4	324	100.0	66	19	AAW64305
5	324	100.0	66	20	AAW39107
6	324	100.0	66	20	AAW38970
7	324	100.0	66	23	AAW29718
8	324	100.0	66	23	AAW17582
9	324	100.0	802	19	AAW81746

10	324	100.0	802	19	AAW64379	Mycobacterium anti
11	324	100.0	802	20	AAW32063	Mycobacterium tube
12	324	100.0	802	20	AAW39224	M. tuberculosis fu
13	324	100.0	802	20	AAW39176	M. tuberculosis fu
14	324	100.0	802	20	AAW39081	M. tuberculosis fu
15	324	100.0	802	20	AAW39033	M. tuberculosis fu
16	324	100.0	802	23	AAU74592	Antigenic fusion p
17	324	100.0	802	23	AAU74592	M. tuberculosis Tb
18	61.5	19.0	253	12	AAW13507	P.denitrificans CO
19	61.5	19.0	368	22	ABW69744	Drosophila melanog
20	61.5	19.0	657	21	AAW57582	Human papillomavir
21	61.5	19.0	657	22	ABW6424	Human papillomavir
22	61.5	19.0	1477	22	ABW61960	Drosophila melanog
23	61	18.8	330	22	AAW38359	Salmonella typhi c
24	60	18.5	224	23	ABP31340	Human ORF313 prote
25	60	18.5	535	24	ABU08348	Cotton (+)-delta-c
26	60	18.5	1624	23	ABU05888	M. tuberculosis an
27	59.5	18.4	1938	24	ABP76679	Streptomyces virid
28	59	18.2	410	23	AAE17876	Sucrose-specific p
29	59	18.2	984	22	ABW61888	Drosophila melanog
30	59	18.2	1221	23	AAE24373	Fruit fly E93 prog
31	58.5	18.1	324	21	AAW39229	Gene 2 human secre
32	58.5	18.1	324	21	AAW39230	Human secreted pro
33	58.5	18.1	398	21	ABW43819	Human cancer assoc
34	58.5	18.1	398	23	ABP41571	Human ovarian anti
35	58.5	18.1	1005	21	AAW31946	Human cytoskeleton
36	58	17.9	110	21	AAW5654	Arabidopsis thalia
37	58	17.9	191	21	AAW24668	Plant SDF encoded
38	58	17.9	617	22	AAW25464	Human mdt protein
39	58	17.9	974	19	AAW55960	Human transient re
40	58	17.9	1737	23	AAO14211	Human transporter
41	58	17.9	1738	23	ABW8491	Human Canlon prote
42	58	17.9	1738	24	ABP96139	Human Canlon chann
43	58	17.9	1738	24	ABP57388	Human BION1 protei
44	57.5	17.7	255	13	AAW21516	Polyfunctional pro
45	57.5	17.7	441	20	AAW40501	A. thaliana yael p

ALIGNMENTS

RESULT 1

AAW32433

ID AAW32433 standard; Protein; 66 AA.

AC AAW32433;

XX

XX 08-JAN-1998 (first entry)

DT

XX Mycobacterium tuberculosis antigen TBra3.

DE

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX

OS Mycobacterium tuberculosis.

XX

XX WO9709428-A2.

XX

XX 13-MAR-1997.

XX

XX 30-AUG-1996; 96WO-US14674.

XX

XX 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659583.

XX (CORI-) CORIXA CORP.

PA Campos-neto A. Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX Twardzik DR, Vedwick TH,

PI

XX

DR WPI; 1997-192903/17.
 DR N-PSDB; AAT91475.
 XX
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 PT
 XX
 PS Example 3; Page 113; 169pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, TbrA3. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 18; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 DB 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 2
 AAW32365
 ID AAW32365 standard; Protein; 66 AA.
 XX
 AC AAW32365;
 XX
 DT 13-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA3.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9709429-A2.
 XX
 PD 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 DR WPI; 1997-192904/17.
 DR N-PSDB; AAT91412.
 XX
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 PS Example 3; Page 123; 190pp; English.

XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, TbrA3. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 66 AA;
 Query Match 100.0%; Score 324; DB 18; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 DB 1 VIDIIGTPTSWEQAAAEAVQARDSDVDIRVARVIEQDMAVDSAGKITRIKLEVSFKM 60
 QY 61 RPAQPR 66
 DB 61 RPAQPR 66
 RESULT 3
 AAW81668
 ID AAW81668 standard; Protein; 66 AA.
 XX
 AC AAW81668;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide TbrA3.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-261042/23.
 DR N-PSDB; AAV64461.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Example 3; Page 109-110; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 XX
 SQ Sequence 66 AA;

Query Match 100.0%; Score 324; DB 19; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60
DB 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60

QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 4

AAW64305
ID AAW64305 standard; Protein; 66 AA.

XX
AC AAW64305;

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbrA3.

XX Tuberculosis; infection; diagnosis; antigen; TbrA3.

XX Mycobacterium tuberculosis strain H37Ra.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX N-PSDB; AAW44353.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis

XX Example 3; Page 114; 250pp; English.

XX This polypeptide comprises Mycobacterium tuberculosis soluble
CC antigen TbrA3. It is encoded by a DNA sequence (see AAW44353)
CC isolated from a M. tuberculosis strain H37Ra expression library
CC with rabbit anti-sera raised against M. tuberculosis supernatant.
CC No significant homology was found between TbrA3 and Genebank
CC database sequences. The invention relates to compositions and
CC methods for diagnosing tuberculosis. It provides polypeptides
CC (see AAW64291-W64379) comprising an antigenic portion of a soluble
CC M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using the
CC above polypeptides, antibodies or oligonucleotide probes and
CC primers, for the diagnosis of tuberculosis.

XX Sequence 66 AA;

Query Match 100.0%; Score 324; DB 19; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60
DB 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 5

AAV39107
ID AAV39107 standard; Protein; 66 AA.

XX
AC AAV39107;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis antigen TbrA3 amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.

XX Mycobacterium tuberculosis.

XX WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

XX Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.

XX N-PSDB; AAZ19263.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions

XX Example 3; Page 107; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAV39083 to
CC AAV39225 are used in the exemplification of the present invention.

XX Sequence 66 AA;

Query Match 100.0%; Score 324; DB 20; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60
DB 1 VIDIIGTSPTEWQAAAEAVQARDSDVDIRVARVIEQDMVDSAGKITRYIKLEVSPKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66

RESULT 6
AA38970
ID AAY38970 standard; Protein; 66 AA.
XX AC
XX AAY38970;
XX AC
DT 05-NOV-1999 (first entry)
XX DT
XX M. tuberculosis recombinant antigen protein Tbra3.
DE DE
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX KW
OS Mycobacterium tuberculosis.
XX OS
PN WO9942118-A2.
XX PN
PD 26-AUG-1999.
XX PD
PF 17-FEB-1999; 99WO-US03265.
XX PF
PR 05-MAY-1998; 98US-0072596.
XX PR
PR 18-FEB-1998; 98US-0024753.
XX PR
XX (CORI-) CORIXA CORP.
PA PA
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX PI
XX WPI; 1999-527416/44.
DR N-PSDB; AA219051.
XX DR
XX New polypeptide comprising antigenic portions of M. tuberculosis
PT PT
PS Example 3; Page 151; 323pp; English.
XX PS
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX CC
SQ Sequence 66 AA;
Query Match 100.0%; Score 324; DB 20; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSTSEWQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSTSEWQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66
RESULT 7
AAE29718
ID AAE29718 standard; Protein; 66 AA.
XX AC
XX AAE29718;
XX AC
DT 27-JAN-2003 (first entry)
XX DT
XX Mycobacterium tuberculosis Tbra3 antigenic protein.
DE DE
KW Vaccine; immunity; diagnostic agent; gene therapy; Tbra3 antigen.
XX KW
XX Mycobacterium tuberculosis.
OS OS

XX WO200272792-A2.
PN PN
XX 19-SEP-2002.
PD PD
XX 13-MAR-2002; 2002WO-US08223.
XX PF
XX 13-MAR-2001; 2001US-275837P.
XX PR
XX (CORI-) CORIXA CORP.
PA PA
XX Skeiky Y, Brannon M, Guderian J;
PI PI
XX WPI; 2002-759844/82.
DR N-PSDB; AAD47095.
XX DR
XX New recombinant nucleic acid molecule comprising a Leishmania TSA,
PT Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective
PT immunity against pathogenic microorganisms e.g. Leishmania and
PT Mycobacterium tuberculosis
XX PT
XX Disclosure; Page 107-108; 155pp; English.
PS PS
XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides,
CC as in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC M. tuberculosis Tbra3 antigenic protein.
XX CC
SQ Sequence 66 AA;
Query Match 100.0%; Score 324; DB 23; Length 66;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSTSEWQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 1 VIDIICTSTSEWQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
QY 61 RPAQPR 66
DB 61 RPAQPR 66
RESULT 8
AAE17582
ID AAE17582 standard; Protein; 66 AA.
XX AC
XX AAE17582;
XX AC
DT 22-APR-2002 (first entry)
XX DT
XX Mycobacterium species Tbra3 protein.
DE DE
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; Tbra3 protein.
XX KW
XX Mycobacterium sp.
OS OS
XX WO200198460-A2.
PN PN
PD 27-DEC-2001.
XX PD
XX 20-JUN-2001; 2001WO-US19959.
XX PF
XX

```
PR 20-JUN-2000; 2000US-0597796.
PR 01-FEB-2001; 2001US-265737P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
DR N-PSDB; AAD28353.
XX
XX Composition comprising MTE39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject
XX
XX Claim 9; Page 123-124; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC TBra3 protein.
XX
XX Query Match 100.0%; Score 324; DB 23; Length 66;
XX Best Local Similarity 100.0%; Pred. No. 1e-35;
XX Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VIDIIGTPTSWEQAAAEAVQRARDSVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
XX |||||
XX 1 VIDIIGTPTSWEQAAAEAVQRARDSVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
XX
XX 61 RPAQPR 66
XX |||||
XX 61 RPAQPR 66
XX
XX RESULT 9
XX AAW81746
XX ID AAW81746 standard; Protein; 802 AA.
XX
XX AAW81746;
XX
XX 27-JAN-1999 (first entry)
XX
XX M. tuberculosis fusion protein Tbf-2.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX vaccine; pharmaceutical; infection; diagnosis.
XX
XX Synthetic.
XX OS Mycobacterium tuberculosis.
XX OS WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX
```

```
PF 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX 11-OCT-1996; 96US-0730510.
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-261042/23.
XX N-PSDB; AAV64567.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and for diagnosis, treatment and prevention of tuberculosis
XX
XX Disclosure; Page 208-211; 230pp; English.
XX
XX This sequence represents the fusion protein Tbf-2 which is composed of
XX immunogenic polypeptides from Mycobacterium tuberculosis (Mt). This
XX protein is used in a method for inducing protective immunity against
XX tuberculosis (TB). This sequence can be formulated into vaccines
XX and/or pharmaceutical compositions for immunising against
XX M. tuberculosis infection or may be used for the diagnosis of TB.
XX
XX Sequence 802 AA;
XX
XX Query Match 100.0%; Score 324; DB 19; Length 802;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-34;
XX Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VIDIIGTPTSWEQAAAEAVQRARDSVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
XX |||||
XX 9 VIDIIGTPTSWEQAAAEAVQRARDSVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
XX
XX 61 RPAQPR 66
XX |||||
XX 69 RPAQPR 74
XX
XX RESULT 10
XX AAW64379
XX ID AAW64379 standard; Protein; 802 AA.
XX
XX AAW64379;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium antigen Tbf2 protein fusion.
XX
XX Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;
XX Tbf-2.
XX
XX Mycobacterium tuberculosis.
XX OS Synthetic.
XX
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX N-PSDB; AAV55801.
XX
```

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX Example 7; Page 223-226; 250pp; English.
XX This polypeptide comprises a fusion protein, designated Tbf-2,
CC composed of Mycobacterium tuberculosis antigens TbrA3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. Tbf-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression
CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.
XX
SQ Sequence 802 AA;
Query Match 100.0%; Score 324; DB 19; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
QY 61 RPAQPR 66
DB 69 RPAQPR 74
RESULT 11
AAV32063
ID AAV32063 standard; Protein; 802 AA.
AC AAV32063;
DT 17-JAN-2000 (first entry)
DE Mycobacterium tuberculosis antigen fusion protein Tbf-2.
XX Tuberculosis; antigen; fusion protein; Tbf-2; TbrA3; 38kD; Tb38-1;
KW DPEP; diagnosis; therapy; vaccine; immunogen.
XX Mycobacterium tuberculosis.
OS WO9951748-A2.
PN 07-APR-1999; 99WO-US07717.
XX 07-APR-1999; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
PA Skeiky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
DR N-PSDB; AAZ20198.
XX New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX Claim 1; Fig 5G-J; 83pp; English.
PS This sequence represents a recombinant Mycobacterium tuberculosis
CC

CC tetra-antigen fusion protein, termed Tbf-2, composed of the antigens
CC TbrA3, 39kD, Tb38-1 and DPEP. The fusion protein is expressed in
CC host cells using a vector carrying a polynucleotide (see AAZ20198)
CC comprising the 4 coding sequences. The invention provides fusion
CC proteins (see AAY32059-71) containing at least 2 M. tuberculosis
CC antigens. The new fusion proteins and polynucleotides encoding
CC them are useful as vaccines for preventing tuberculosis (claimed),
CC for diagnosis (via in vitro assays or intradermal skin tests for
CC detection of anti-M. tuberculosis antibodies), monitoring of
CC disease progression, and treatment of tuberculosis. They are more
CC effective immunogens than mixtures of the individual protein
CC components.
XX
SQ Sequence 802 AA;
Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 68
QY 61 RPAQPR 66
DB 69 RPAQPR 74
RESULT 12
AAV39224
ID AAV39224 standard; Protein; 802 AA.
XX AAV39224;
AC AAV39224;
DT 05-NOV-1999 (first entry)
DE M. tuberculosis fusion protein Tbf-6 amino acid sequence.
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX Synthetic.
OS Mycobacterium tuberculosis.
XX WO9942076-A2.
PN 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03268.
XX 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
DR N-PSDB; AAZ19457.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX Claim 37; Page 271-273; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to

CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 13
AAY39176
ID AAY39176 standard; Protein; 802 AA.
XX
AC AAY39176;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis fusion protein TbF-2 amino acid sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Synthetic.
CS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03268.
XX
PR 05-MAY-1998; 98US-0072967.
PR 18-FEB-1998; 98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.
DR N-PSDB; AA219368.
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
PS Disclosure; Page 205-208; 299pp; English.

CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 14
AAY39081
ID AAY39081 standard; Protein; 802 AA.
XX
AC AAY39081;
XX
DT 05-NOV-1999 (first entry)
XX
DE M tuberculosis fusion protein TbF-6.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942118-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US03265.
XX
PR 05-MAY-1998; 98US-0072596.
PR 18-FEB-1998; 98US-0024753.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527416/44.
DR N-PSDB; AA219245.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis
PS Example 10; Page 316-318; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
SQ Sequence 802 AA;

Query Match 100.0%; Score 324; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.4e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
DB 9 VIDIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
DB 69 RPAQPR 74

RESULT 15
 AAY39033
 ID AAY39033 standard; Protein; 802 AA.
 XX
 XX AAY39033;
 AC
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbF-2.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 XX 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AAZ19156.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 10; Page 251-253; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 802 AA;

Query Match	100.0%	Score 324;	DB 20;	Length 802;
Best Local Similarity	100.0%	Pred. No. 2.4e-34;		
Matches 66;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	VIDIIGTPTSEQAARQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM	60
Db	9	VIDIIGTPTSEQAARQARDSDVDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM	60
QY	61	RPAQPR	56
Db	69	RPAQPR	74

Search completed: November 21, 2003, 16:03:12
 Job time : 4.99592 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 2.57415 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIGSTSTNEQRAAEAV.....KITYRIKLEVSFKMRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559496 residues
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA: *
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep: *
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep: *
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep: *
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep: *
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep: *
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep: *
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep: *
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep: *
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep: *
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep: *
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep: *
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep: *
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep: *
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep: *
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep: *
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep: *
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep: *
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#				SUMMARIES	
Result No.	Score	Query Length	ID	Description	
1	324	100.0	66	12 US-10-084-843-77	Sequence 77, Appl
2	324	100.0	66	12 US-10-193-002-78	Sequence 78, Appl
3	324	100.0	66	12 US-10-098-7328-37	Sequence 37, Appl
4	324	100.0	802	9 US-03-287-849-10	Sequence 10, Appl
5	324	100.0	802	12 US-10-084-843-214	Sequence 214, Appl
6	324	100.0	802	12 US-10-084-843-351	Sequence 351, Appl
7	324	100.0	802	12 US-10-193-002-209	Sequence 209, Appl
8	324	100.0	802	12 US-10-193-002-346	Sequence 346, Appl
9	324	100.0	802	12 US-10-359-460-10	Sequence 10, Appl
10	63.5	19.6	342	12 US-10-166-225A-63	Sequence 63, Appl
11	62.5	19.3	561	15 US-10-078-770-154	Sequence 154, Appl
12	61.5	19.0	1073	16 US-10-210-115-20	Sequence 20, Appl
13	61	18.8	330	9 US-09-815-242-13952	Sequence 13952, A
14	60	18.5	535	14 US-10-067-534-2	Sequence 2, Appl
15	60	18.5	1624	16 US-10-080-170-539	Sequence 539, Appl

16	59	18.2	410	10	US-09-905-176-26	Sequence 26, Appl
17	59	18.2	1221	14	US-10-016-768-11	Sequence 11, Appl
18	58.5	18.1	398	9	US-09-925-301-1264	Sequence 1264, Appl
19	58	17.9	828	15	US-10-198-070-98	Sequence 98, Appl
20	58	17.9	836	15	US-10-198-070-97	Sequence 97, Appl
21	58	17.9	974	12	US-10-153-244-106	Sequence 106, Appl
22	58	17.9	974	12	US-10-210-152-23	Sequence 23, Appl
23	58	17.9	977	15	US-10-198-070-108	Sequence 108, Appl
24	58	17.9	982	15	US-10-198-070-63	Sequence 63, Appl
25	58	17.9	982	15	US-10-198-070-72	Sequence 72, Appl
26	58	17.9	1738	15	US-10-178-791-1	Sequence 1, Appl
27	58	17.9	1738	15	US-10-178-791-3	Sequence 3, Appl
28	57.5	17.7	500	10	US-09-981-947A-7	Sequence 7, Appl
29	57.5	17.7	861	15	US-10-156-761-11328	Sequence 11328, A
30	57	17.6	323	15	US-10-198-070-64	Sequence 64, Appl
31	56.5	17.4	1115	10	US-09-866-582-40	Sequence 40, Appl
32	56	17.3	94	12	US-10-029-386-28176	Sequence 28176, A
33	56	17.3	515	12	US-10-306-905-12	Sequence 12, Appl
34	56	17.3	2548	9	US-09-851-682A-1	Sequence 1, Appl
35	55.5	17.1	460	16	US-10-278-536-2	Sequence 2, Appl
36	55.5	17.1	489	10	US-09-826-581-6	Sequence 6, Appl
37	55.5	17.1	613	12	US-10-120-145-5	Sequence 5, Appl
38	55.5	17.1	722	15	US-10-136-841-6	Sequence 6, Appl
39	55	17.0	85	15	US-10-156-761-14857	Sequence 14857, A
40	55	17.0	888	15	US-10-097-340-161	Sequence 161, Appl
41	55	17.0	928	9	US-09-801-574-44	Sequence 44, Appl
42	54.5	16.8	310	15	US-10-156-761-12054	Sequence 12054, A
43	54.5	16.8	1107	15	US-10-153-219-4	Sequence 4, Appl
44	54.5	16.8	1143	15	US-10-153-219-6	Sequence 6, Appl
45	54.5	16.8	1143	15	US-10-153-219-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-10-084-843-77
; Sequence 77, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

```
/
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C9
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 77:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 66 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-084-843-77

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 2
US-10-193-002-78
/ Sequence 78, Application US/10193002
/ Publication No. US20030135026A1
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonia
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Hendrickson, Ronald C.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
/ TUBERCULOSIS
/ NUMBER OF SEQUENCES: 350
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/193,002
/ FILING DATE: 10-Jul-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/072,596
/ FILING DATE: 05-MAY-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.417C9
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 78:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 66 amino acids
```

```
/
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-193-002-78

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 3
US-10-098-732A-37
/ Sequence 37, Application US/10098732A
/ Publication No. US20030175294A1
/ GENERAL INFORMATION:
/ APPLICANT: Skeiky, Yasir
/ APPLICANT: Brannon, Mark
/ APPLICANT: Guderian, Jeffrey
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
/ TITLE OF INVENTION: Leishmania Antigen
/ FILE REFERENCE: 014058-012010US
/ CURRENT APPLICATION NUMBER: US/10/098,732A
/ CURRENT FILING DATE: 2003-04-29
/ PRIOR APPLICATION NUMBER: US 60/275,837
/ PRIOR FILING DATE: 2001-03-13
/ NUMBER OF SEQ ID NOS: 80
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 37
/ LENGTH: 66
/ TYPE: PRT
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: Tbra3
US-10-098-732A-37

Query Match 100.0%; Score 324; DB 12; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.7e-35;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60
Db 1 VIDIIGTSPTEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITIRIKLEVSFKM 60

QY 61 RPAQPR 66
Db 61 RPAQPR 66

RESULT 4
US-09-287-849-10
/ Sequence 10, Application US/09287849
/ Patent No. US20020009459A1
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Alderson, Mark
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
/ TITLE OF INVENTION: and Their Uses
/ FILE REFERENCE: 014058-009020US
/ CURRENT APPLICATION NUMBER: US/09/287,849
```

; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10

Query Match 100.0%; Score 324; DB 9; Length 802;
Best Local Similarity 100.0%; Pred.No.3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSPTSWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 5
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-10-084-843-214

Query Match 100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred.No.3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSPTSWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSPTSWEQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 6
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 351:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 7
US-10-193-002-209
; Sequence 209, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-10-193-002-209

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
Db 69 RPAQPR 74

RESULT 8
US-10-193-002-346
; Sequence 346, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 60
Db 9 VIDIIGTSTSEWQAAAEAVQARSDVDDIRVARVIEQDMVDSAGKITRYIKLEVSFKM 68

QY 61 RPAQPR 66
```

```
Db          |||||
69 RPAQPR 74

RESULT 9
US-10-359-460-10
; Sequence 10, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dallon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tetra-fusion
US-10-359-460-10

Query Match          100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
|||
69 RPAQPR 74

RESULT 10
US-10-166-225A-63
; Sequence 63, Application US/10166225A
; Publication No. US20030148416A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMSELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISSEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/10/166,225A
; CURRENT FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 63

Query Match          100.0%; Score 324; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.8e-34;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIDIIGTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 60
Db 9 VIDIIGTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITVRIKLEVSFKM 68

QY 61 RPAQPR 66
|||
69 RPAQPR 74

RESULT 11
US-10-078-770-154
; Sequence 154, Application US/10078770
; Publication No. US20030003471A1
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Forge, Charlie
; APPLICANT: Miao, Guo-Hua
; TITLE OF INVENTION: cDNAs Encoding Polypeptides
; FILE REFERENCE: BB-1365 US NA
; CURRENT APPLICATION NUMBER: US/10/078,770
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/614,188
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,400
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/153,534
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/161,223
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/159,878
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: 60/157,401
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/143,419
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,409
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 154
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Zea mays
US-10-078-770-154

Query Match          19.3%; Score 62.5; DB 15; Length 561;
Best Local Similarity 32.1%; Pred. No. 12;
Matches 18; Conservative 12; Mismatches 23; Indels 3; Gaps 2;

QY 14 QAAAEAVQRAEDSVDDIRVARVIEQDM--AVDSAGKITVRI-KLEVSFKMREPAQPR 66
Db 228 EAAKVLVKEKGDVVAQLKAAKASKOEISTAVDELNRAKEIVSKLEERFKLKGIPR 283

RESULT 12
US-10-210-115-20
; Sequence 20, Application US/10210115
; Publication No. US20030129708A1
; GENERAL INFORMATION:
; APPLICANT: Puitsyn, Leonid R
; APPLICANT: Smirnov, Sergey V
; APPLICANT: Altman, Irina B
; APPLICANT: No. US20030129708A1Ikovva, Anna E
; APPLICANT: Kotliarova, Veronika A
```


Search completed: November 21, 2003, 16:38:11
Job time : 2.57415 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 1.37687 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIICTSPTSWEQAAAEAV.....KITYRIKLEVSFWMRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	71	2 F70833	probable protein t
2	101	31.2	77	2 E84298	hypothetical prote
3	69	21.3	71	2 E83641	hypothetical prote
4	66	20.4	69	2 D95850	hypothetical prote
5	65.5	20.2	1073	2 D90633	carbamoyl-phosphat
6	65.5	20.2	1073	2 C85494	carbamoyl-phosphat
7	64	19.8	96	2 AE1822	transcription regu
8	63.5	19.6	342	2 C71704	hypothetical prote
9	63.5	19.6	342	2 H97792	carotenoid biosynt
10	62	19.1	282	2 AD3528	flagellin (importe
11	62	19.1	1094	2 S49313	protein kinase - s
12	61.5	19.0	253	2 H36145	El protein - Pse
13	61.5	19.0	657	1 W1WL18	beta-glucosidase (
14	61.5	19.0	947	2 S08243	carbamoyl-phosphat
15	61.5	19.0	1073	1 SYECP	tumor supressor pr
16	61.5	19.0	1477	2 T13797	hypothetical prote
17	61	18.8	94	2 E84227	myosin-RhoGAP prot
18	61	18.8	2626	2 T31099	carbamoyl-phosphat
19	60.5	18.7	1075	2 A30510	hypothetical prote
20	60	18.5	181	2 S23604	cell division prot
21	60	18.5	246	2 AF2652	hypothetical prote
22	60	18.5	267	2 C97434	ABC transporter (A
23	60	18.5	396	2 H59803	Probable Helix-tur
24	60	18.5	1624	2 C70867	Desmin - African c
25	59.5	18.4	458	2 A43554	ABC transporter, A
26	59.5	18.4	620	2 H69382	probable leucyl-tr
27	59.5	18.4	959	2 H72699	acetyl-CoA acetyl
28	59	18.2	393	2 AH2916	chain A, unligande
29	59	18.2	393	2 C97691	

ALIGNMENTS

RESULT 1

F70833
Probable protein transport protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70833
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295387; PMID:9634230
A:Accession: F70833
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-71 <COL>
A:Cross-references: GB:AL021931; GB:AL123456; NID:G3261526; PIDN:CAA17385.1; PID:e1252494
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: sec

Query Match 100.0%; Score 324; DB 2; Length 71;
Best Local Similarity 100.0%; Pred. No. 5e-30;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIICTSPTSWEQAAAEAVQARDSDIRVARVIEQDMAVDSAGKITRYIKLEVSFWM 60
DB 6 VIDIICTSPTSWEQAAAEAVQARDSDIRVARVIEQDMAVDSAGKITRYIKLEVSFWM 65
QY 61 RPAQPR 66
DB 66 RPAQPR 71

RESULT 2

E84298
hypothetical protein Vng1446h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84298
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabloner, J.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84298
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-77 <STO>

A;Accession: S52076
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1094 <WE2>
A;Cross-references: EMBL:Z37981; NID:G551445; PIDN:CAA86053.1; PID:G551446
C;Genetics:
A;Introns: 35/3; 104/1; 166/2
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
F:20-281/Domain: protein kinase homology <KIN>
F:28-36/Region: protein kinase ATP-binding motif

Query Match 19.1%; Score 62; DB 2; Length 1094;
Best Local Similarity 30.0%; Pred. No. 68;
Matches 15; Conservative 13; Mismatches 20; Indels 2; Gaps 1;

QY 13 EQAAAEAVQARDSDVDIRVARVI--EODMAVDSAGKITRYIKLEVSFKM 60
DB 893 EVKAEPTKAESEVEDVKVEPIKVEPVKVEEPIKVEPVKVEEPIKV 942

RESULT 12
H36145
cobM protein - Pseudomonas sp.
C;Species: Pseudomonas sp.
C;Date: 30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change 24-Sep-1999
C;Accession: H36145
J. Bacteriol. 172, 5980-5990, 1990
R. Crouzet, J.; Cameron, B.; Cauchots, L.; Rigault, S.; Rouyez, M.C.; Blanche, F.; Thibaut
A;Title: Genetic and sequence analysis of an 8.7-kilobase Pseudomonas denitrificans frag
A;Reference number: A36145; MUID:91008976; PMID:2211521
A;Accession: H36145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <CRO>
A;Cross-references: GB:M59301; GB:M32224; NID:G151170; PIDN:AAA25801.1; PID:G151178
A;Note: the source is designated as Pseudomonas denitrificans
C;Superfamily: precorrin-3 methylase

Query Match 19.0%; Score 61.5; DB 2; Length 253;
Best Local Similarity 32.3%; Pred. No. 16;
Matches 21; Conservative 14; Mismatches 23; Indels 7; Gaps 3;

QY 4 IIGTPTGWEQAAAEAVQARDSDVDIRVARVIEODMAVDSAGKITRYIKLE---VSFKM 60
DB 51 IVDTAPMSLDELEAVYVAEAEGLD---VARLHSGDISVWSAVAEQIR-RLEKHGIATM 106

QY 61 RPAQP 65
DB 107 TPGVP 111

RESULT 13
EWL18
EI protein - human papillomavirus type 18
C;Species: human papillomavirus type 18
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C;Accession: C26165; C26251
R. Seedorf, K.; Oltersdorf, T.; Kraemer, G.; Roewekamp, W.
EMBO J. 6, 139-144, 1987
A;Title: Identification of early proteins of the human papilloma viruses type 16 (HPV 16
A;Reference number: A91068; MUID:87218459; PMID:3034571
A;Accession: C26165
A;Molecule type: DNA
A;Residues: 1-274 <SE>
A;Cross-references: GB:X04773; NID:G60876; PIDN:CAA28468.1; PID:G60879
R. Cole, S.T.; Panos, O.
J. Mol. Biol. 193, 599-608, 1987
A;Title: Nucleotide sequence and comparative analysis of the human papillomavirus type 1
A;Reference number: A92937; MUID:87283882; PMID:3039146
A;Accession: C26251
A;Molecule type: DNA
A;Residues: 1-657 <COL>

A;Cross-references: GB:X05015; NID:G60975; PIDN:CAA28666.1; PID:G60978
C;Superfamily: papillomavirus E1 protein
C;Keywords: early protein

Query Match 19.0%; Score 61.5; DB 1; Length 657;
Best Local Similarity 29.2%; Pred. No. 44;
Matches 19; Conservative 14; Mismatches 29; Indels 3; Gaps 1;

QY 1 VIDIIGTSPTSWEQA--AAEAVQARDSDVDIRVARVIEODMAVDSAGKITRYIKLEVS 57
DB 47 WVDFTDTQTTFCEQAELETAQALFRAQEVHNDQVHLVKRKFAGGSTENSPIGSERLEVD 106

QY 58 FKMRP 62
DB 107 TELSP 111

RESULT 14
S08243
beta-glucosidase (EC 3.2.1.21) - Ruminococcus albus
C;Species: Ruminococcus albus
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C;Accession: S08243
R. Ohmura, K.; Takano, M.; Shimizu, S.
Nucleic Acids Res. 18, 671, 1990
A;Title: DNA sequence of a beta-glucosidase from Ruminococcus albus.
A;Reference number: S08243; MUID:90175009; PMID:2106673
A;Accession: S08243
A;Molecule type: DNA
A;Residues: 1-947 <OHM>
A;Cross-references: EMBL:X15415; NID:G45967; PIDN:CAA33461.1; PID:G45968
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 19.0%; Score 61.5; DB 2; Length 947;
Best Local Similarity 33.3%; Pred. No. 66;
Matches 18; Conservative 11; Mismatches 22; Indels 3; Gaps 2;

QY 13 EQAAAEAVQARDSDVDIRVARV--IEQDMAVDSAGKITRYIKLEVSFQMRQAQ 64
DB 774 EDEAVEINKPAETVDDGEGDRVFLDGLDITDMSGVKTER-NLDYSFTVDVAQ 826

RESULT 15
SYECCP
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [validated]
N;Alternate names: carbamoyl-phosphate synthetase (glutamine-hydrolyzing) ammonia chain
C;Species: Escherichia coli
C;Date: 19-Feb-1984 #sequence_revision 31-Dec-1991 #text_change 01-Mar-2002
C;Accession: A01198; S40556; A64724
R. Nyunoya, H.; Lusty, C.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 4629-4633, 1983
A;Title: The carb gene of Escherichia coli: a duplicated gene coding for the large subunit
A;Reference number: A01198; MUID:83273869; PMID:8308632
A;Accession: A01198
A;Molecule type: DNA
A;Residues: 1-1073 <NYU>
A;Cross-references: GB:J01597; NID:G145461; PIDN:AAA23539.1; PID:G145464
R. Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobuc
Submitted to the EMBL Data Library, December 1992
A;Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A;Reference number: S40531
A;Accession: S40556
A;Molecule type: DNA
A;Residues: 1-1073 <YUR>
A;Cross-references: EMBL:D10483; NID:G216434; PIDN:BA01311.1; PID:G216460
R. Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64724
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-1073 <BLAST>
A;Cross-references: GB:AE000113; GB:U00096; NID:g23s7095; PTDN:AAC73144.1; PTD:g1786216;
A;Experimental source: strain K-12, substrain NG1655

```

Query Match      19.0%; Score 61.5; DB 1; Length 1073;
Best Local Similarity 29.0%; Pred.No.76;
Matches 18; Conservative 18; Mismatches 19; Indels 7; Gaps 2;

QY      2 IDIGTSWEQAAEAQVARDSDVDDIRV-----ARVIEQDMVDSAGKITRYIKLEV 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
db 659 VPVIGTSPOAIDR--AEDRRFQHAVERLKKOPANATVTAIEMAEKAEIGYPLVPR 716
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: November 21, 2003, 16:09:52
Job time : 2.37687 secs

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 0.778231 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-6

Perfect score: 324

Sequence: 1 VIDIIIGTSPTEQAAAEV.....KITYRIKLEVSFQRPQAPR 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	65.5	20.2	1	CARB_ECO57
2	63.5	19.6	342	1 ID12_RICGN
3	63.5	19.6	342	1 ID12_RICPR
4	63	19.4	731	1 HUTU_CABEL
5	61.5	19.0	253	1 COBM_PSEDE
6	61.5	19.0	657	1 VBI_HPV18
7	61.5	19.0	947	1 BGLS_RUMAL
8	61.5	19.0	1072	1 CARB_ECOLI
9	61	18.8	191	1 RS7_BRAOL
10	60.5	18.7	351	1 ID12_RHILQ
11	60.5	18.7	1074	1 CARB_SALTI
12	60.5	18.7	1074	1 CARB_SALTY
13	60	18.5	321	1 KGFY_PASNU
14	59.5	18.4	458	1 DESM_XENLA
15	59.5	18.4	959	1 SYL_AERPE
16	59.5	18.4	988	1 E4L2_MOUSE
17	59	18.2	505	1 SCRY_SALTY
18	58.5	18.1	361	1 COBT_MYCTU
19	58.5	18.1	1005	1 E4L2_HUMAN
20	58	17.9	974	1 TP4_MOUSE
21	58	17.9	977	1 TRP4_HUMAN
22	58	17.9	977	1 TRP4_RAT
23	58	17.9	981	1 TRP4_BOVIN
24	58	17.9	1577	1 MYSH_ACACA
25	57.5	17.7	286	1 PANB_RHILQ
26	57.5	17.7	501	1 GRIS_HUMAN
27	57.5	17.7	1078	1 CARB_BUCAL
28	57	17.6	1184	1 DP3A_MYCTU
29	56.5	17.4	355	1 VPI_HAEIN
30	56.5	17.4	1332	1 YSVI_CABEL
31	56	17.3	473	1 SVC_METMA
32	56	17.3	515	1 LBUI_HAEIN
33	56	17.3	754	1 ASFH_BOVIN

RESULT 1				
CARB_ECO57				
ID	CARB_ECO57	STANDARD;	PRT;	1072 AA.
AC	Q8XA38;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).			
DE	Phosphate synthetase ammonia chain).			
GN	CARB OR Z0038 OR ECS0036 OR SF0030.			
OS	Escherichia coli O157:H7, and			
OS	Shigella flexneri.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=83334, 623;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.F., Potamouis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RL	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL	Nature 409:529-533(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=E.coli; STRAIN=O157:H7 / RMD 0509952;			
RX	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT	O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL	DNA Res. 8:111-22(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;			
RX	MEDLINE=2272406; PubMed=12384590;			
RA	Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,			
RA	Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,			
RA	Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,			
RA	Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,			
RA	Yu J.;			
RT	"Genome sequence of Shigella flexneri 2a: insights into pathogenicity			
RT	through comparison with genomes of Escherichia coli K12 and O157.;"			
RL	Nucleic Acids Res. 30:4432-4441(2002).			
CC	-I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +			
CC	phosphate + L-glutamate + carbamoyl phosphate.			
CC	-I- COFACTOR: Binds 3 manganese ions per subunit (By similarity).			
CC	-I- PATHWAY: Arginine biosynthesis.			
CC	-I- PATHWAY: Pyrimidine biosynthesis; first step.			
CC	-I- SUBUNIT: Composed of two chains; the small (or glutamine) chain			

Q8fpa8 corynebacte
Q8fnp5 corynebacte
P33593 escherichia
Q9ugi9 homo sapien
P08236 homo sapien
P13332 bacterioph
O35451 mus musculu
P48612 drosophila
P46700 mycobacteri
P71017 bacillus su
P21134 aphanocapsa
Q9hmt9 halobacteri

ALIGNMENTS

CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC Tetramer of heterodimers (alpha,beta)4 (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; A005180; AAG54335.1; -
 CC EMBL; A002550; BAB33459.1; -
 CC EMBL; A015040; AAN41696.1; -
 CC PIR; C85484; C85484.
 CC PIR; D90633; D90633.
 CC HAMAP; MF_01210; -; 1.
 CC InterPro; IPR006275; Cara_L_glu.
 CC InterPro; IPR005483; CPase_L.
 CC InterPro; IPR005479; CPase_L_D2.
 CC InterPro; IPR005480; CPase_L_D3.
 CC InterPro; IPR005481; CPase_L_N.
 CC InterPro; IPR004382; MGS_L-like.
 CC Pfam; PF00289; CPase_L_chain; 2.
 CC Pfam; PF02786; CPase_L_D2; 2.
 CC Pfam; PF02787; CPase_L_D3; 2.
 CC Pfam; PF02142; MGS; 1.
 CC PRINTS; PR00098; CPASE.
 CC TIGRfam; TIGR01369; CPaseII_lrg; 1.
 CC PROSITE; PS00866; CPASE_1; 2.
 CC PROSITE; PS00867; CPASE_2; 2.
 CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT INIT MET 0
 FT DOMAIN 1 402
 FT DOMAIN 403 552
 FT DOMAIN 553 935
 FT DOMAIN 936 1072
 FT REPEAT 1 552
 FT REPEAT 553 1072
 FT NP_BIND 152 209
 FT NP_BIND 302 353
 FT METAL 284 284
 FT METAL 288 288
 FT METAL 300 300
 FT METAL 828 828
 FT METAL 840 840
 SQ SEQUENCE 1072 AA; 117724 MW; D98867964E058E37 CRC64;
 Query Match 20.2%; Score 65.5; DB 1; Length 1072;
 Best Local Similarity 30.6%; Pred. No. 20;
 Matches 19; Conservative 17; Mismatches 19; Indels 7; Gaps 2;
 QY 2 IDIIIGTSPTSEQAAAEAVQARDSDVIR-----ARVIEQDMVADPSAGKITRYIKLEV 56
 Db 658 VPIVIGTSPTDAIR--AEDRRFQHAVDLKQKQANATVTAIEAVKAKKIGYPLVVRP 715
 QY 57 SF 58
 Db 716 SY 717
 RESULT 2
 IDI2_RICCN STANDARD; PRT; 342 AA.
 AC Q92HW7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE (Isopentenyl pyrophosphate isomerase).
 FNI OR RP452.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

FNI OR RC0744.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 EX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesco-Audiffren P., Fournier P.-E., Barbe V.,
 RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Rault D.;
 RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 CC -!- FUNCTION: Catalyzes the 1,3-allylic rearrangement of the
 CC homoallylic substrate isopentenyl (IPP) to its allylic isomer,
 CC dimethylallyl diphosphate (DMAPP) (By similarity).
 CC -!- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
 CC diphosphate.
 CC -!- COFACTOR: FMN and NADPH (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the IPP isomerase type 2 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE008632; AAL03282.1; -
 CC PIR; H97792; H97792.
 CC HAMAP; MF_00354; -; 1.
 CC InterPro; IPR003009; FMN enzyme.
 KW Isomerase; Isoprene biosynthesis; Flavoprotein; FMN; NADP;
 KW Complete proteome.
 SQ SEQUENCE 342 AA; 37336 MW; D346003CCD99B7A3 CRC64;
 Query Match 19.6%; Score 63.5; DB 1; Length 342;
 Best Local Similarity 25.8%; Pred. No. 11;
 Matches 16; Conservative 13; Mismatches 14; Indels 19; Gaps 2;
 QY 1 VDIIGTSPTSEQAAAEAVQARDSDVIR-----VDIRVARVIEQDMVADVS 44
 Db 215 VLDIAGSGGTSSQVEA---YRATNSLQNRASSFTNWGIPITLDSLKMVREYSKDIPIIT 271
 QY 45 AG 46
 Db 272 SG 273
 RESULT 3
 IDI2_RICPR STANDARD; PRT; 342 AA.
 AC Q9ZD30;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE (Isopentenyl pyrophosphate isomerase).
 FNI OR RP452.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;


```
DR PROSITE; P800840; SUMT 2; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase.
SQ SEQUENCE 253 AA; 26862 MW; 50F5E6F5F67FF733 CRC64;

Query Match
Best Local Similarity 19.0%; Score 61.5; DB 1; Length 253;
Matches 21; Conservative 14; Mismatches 23; Indels 7; Gaps 3;

QY 4 IIGTSPTEQAAAVQARSDVDIRVARVIEQDMVDSAGKITYRIKLE---VSPFM 60
Db 51 IVDTAPMSLDEIAEVVKAEGLD---VARLHSGDLSWSAVAEQIR-RIEKGIATYM 106

QY 61 RPAQP 65
Db 107 TFGVP 111

RESULT 6
VE1_HPV18
ID VE1_HPV18 STANDARD; PRT; 657 AA.
AC P06789;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Replication protein E1.
GN E1.
OS Human papillomavirus type 18.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10582;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283882; PubMed=3039146;
RA Cole S.T., Dancos O.;
RT "Nucleotide sequence and comparative analysis of the human
RT papillomavirus type 18 genome. Phylogeny of papillomaviruses and
RT repeated structure of the E6 and E7 gene products.";
RL J. Mol. Biol. 193:599-608(1987).
RN [2]
RP SEQUENCE OF 1-528 FROM N.A.
RX MEDLINE=88188247; PubMed=2833614;
RA Inagaki Y., Tsunokawa Y., Takebe N., Nawa H., Nakanishi S.,
RA Terada M., Sugimura T.;
RT "Nucleotide sequences of cDNAs for human papillomavirus type 18
RT transcripts in HeLa cells.";
RL J. Virol. 62:1640-1646(1988).
RN [3]
RP SEQUENCE OF 1-30 FROM N.A.
RX MEDLINE=87053870; PubMed=3023067;
RA Schneider-Gaedick A., Schwarz E.;
RT "Different human cervical carcinoma cell lines show similar
RT transcription patterns of human papillomavirus type 18 early genes.";
RL EMBO J. 5:2285-2292(1986).
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05015; CAA28666.1; -
CC EMBL; M20325; AAA99516.1; -
CC EMBL; M26798; AAA46948.1; -
CC PIR; C26165; W1W118.
CC InterPro; IPR001177; Papillom_E1.

DR PROSITE; P800840; SUMT 2; 1.
DR Pfam; PF00519; E1; 1.
DR Pfam; PF00524; E1_N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
KW Nuclear protein.
FT NP_BIND 484 491 ATP (POTENTIAL).
FT CONFLICT 94 94 T -> K (IN REF. 2).
SQ SEQUENCE 657 AA; 73736 MW; B1B568BCE1DB930A CRC64;

Query Match
Best Local Similarity 19.0%; Score 61.5; DB 1; Length 657;
Matches 19; Conservative 14; Mismatches 29; Indels 3; Gaps 1;

QY 1 VIDIIPTSPTEQAA---AAEAVQARSDVDIRVARVIEQDMVDSAGKITYRIKLEVS 57
Db 47 MVDIFDTQGTFCBQAELETAQALPHAQVHNDQAQLVHLKRRKFAAGSGSTENSLGERLEVD 106

QY 58 FKWRP 62
Db 107 TELSP 111

RESULT 7
BGLS_RUMAL
ID BGLS_RUMAL STANDARD; PRT; 947 AA.
AC P15835;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-glucosidase (EC 3.2.1.21) (Gentiobiase) (Cellobiase) (Beta-D-
DE Glucoside glucosylhydrolase).
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=F-40;
RX MEDLINE=90175009; PubMed=2106673;
RA Ohmiya K., Takano M., Shimizu S.;
RT "DNA sequence of a beta-glucosidase from Ruminococcus albus.";
RL Nucleic Acids Res. 18:671-671(1990).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC glucose residues with release of beta-D-glucose.
CC -!- PATHWAY: Cellulose degradation.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15415; CAA33461.1; -
CC PIR; S08243; S08243.
DR InterPro; IPR002772; Glyco_hydro_3C.
DR InterPro; IPR001764; Glyco_hydro_3N.
DR Pfam; PF00933; Glyco_hydro_3_1.
DR Pfam; PF01915; Glyco_hydro_3_C; 1.
DR PRINTS; PR00133; GLHYDRLASE3.
DR PROSITE; PS00775; GLYCOSYL HYDROL_F3; 1.
KW Hydrolase; Glycosidase; Cellulose degradation.
FT ACT_SITE 696 696 BY SIMILARITY.
SQ SEQUENCE 947 AA; 104277 MW; C43B8CDD9D60A115 CRC64;

Query Match
Best Local Similarity 19.0%; Score 61.5; DB 1; Length 947;
Matches 18; Conservative 11; Mismatches 22; Indels 3; Gaps 2;

QY 13 EQAAAEAVQARSDVDIRVARV---IEQDMVDSAGKITYRIKLEVSFKMRPAQ 64
Db 774 EDEAVEVINKPAETVDDGSGDRVFLDDGLDTIDMSGVKTER-NLDYSFTVDVAQ 826
```

RESULT 8

ID CARB_ECOLI STANDARD; PRF; 1072 AA.
 AC P00968;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 GN CARB OR PYRA OR B0033.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 1-1.
 RC STRAIN=K12;
 RX MEDLINE=83273669; PubMed=6308632;
 RA Nyunoya H., Lusty C.J.;
 RT "The carb gene of Escherichia coli: a duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4629-4633(1983).
 RN [2]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=94248073; PubMed=6377309;
 RA Bouvier J., Patte J.-C., Stragier P.;
 RT "Multiple regulatory signals in the control region of the Escherichia coli carA operon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4139-4143(1984).
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N., Isono K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [4]
 RC SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RC SEQUENCE OF 1-6 FROM N.A.
 RX MEDLINE=84248072; PubMed=6330744;
 RA Piette J., Nyunoya H., Lusty C.J., Cunin R., Weyens G., Crabeel M., Charlier D.R.M., Glansdorff N., Piarard A.;
 RT "DNA sequence of the carA gene and the control region of carA: tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in Escherichia coli K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4134-4138(1984).
 RN [6]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=97317071; PubMed=9174345;
 RA Thoden J.B., Holden H.M., Wesenberg G., Raushel F.M., Rayment I.;
 RT "Structure of carbamoyl phosphate synthetase: a journey of 96 A from substrate to product.";
 RL Biochemistry 36:6305-6316(1997).
 RN [7]
 RC X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=98301363; PubMed=9636022;
 RA Thoden J.B., Miran S.G., Phillips J.C., Howard A.J., Raushel F.M., Holden H.M.;
 RT "Carbamoyl phosphate synthetase: caught in the act of glutamine

hydrolysis.";
 RL Biochemistry 37:8825-8831(1998).
 RN [8]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=99190825; PubMed=10089390;
 RA Thoden J.B., Raushel F.M., Benning M.M., Rayment I., Holden H.M.;
 RT "The structure of carbamoyl phosphate synthetase determined to 2.1-A resolution.";
 RL Acta Crystallogr. D 55:8-24(1999).
 RN [9]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=99152507; PubMed=10029528;
 RA Thoden J.B., Wesenberg G., Raushel F.M., Holden H.M.;
 RT "Carbamoyl phosphate synthetase: closure of the B-domain as a result of nucleotide binding.";
 RL Biochemistry 38:2347-2357(1999).
 RN [10]
 RC X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20056035; PubMed=10587438;
 RA Thoden J.B., Huang X., Raushel F.M., Holden H.M.;
 RT "The small subunit of carbamoyl phosphate synthetase: snapshots along the reaction pathway.";
 RL Biochemistry 38:16159-16166(1999).
 RN [11]
 RC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=99357782; PubMed=10428826;
 RA Thoden J.B., Raushel F.M., Wesenberg G., Holden H.M.;
 RT "The binding of inosine monophosphate to Escherichia coli carbamoyl phosphate synthetase.";
 RL J. Biol. Chem. 274:22502-22507(1999).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit.
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC Tetramer of heterodimers (alpha,beta)4.
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V01500; CA24744.1; -.
 DR EMBL; D10483; BAB96602.1; -.
 DR EMBL; AE000113; AAC73144.1; -.
 DR PIR; A01198; SYECCP.
 DR PDB; 1JDB; 17-JUN-98.
 DR PDB; 1A9X; 21-OCT-98.
 DR PDB; 1BX8; 20-APR-99.
 DR PDB; 1CB8; 26-JUL-99.
 DR PDB; 1C30; 10-DEC-99.
 DR PDB; 1C30; 10-DEC-99.
 DR PDB; 1CS0; 10-DEC-99.
 DR PDB; 1KEE; 21-DEC-01.
 DR PDB; 1M6V; 13-NOV-02.
 DR ECODBASE; E133.0; 6TH EDITION.
 DR Ecdgene; EGI0135; carB.
 DR HAMAP; MF_01210; -; 1.
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF02889; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.

DR Pfam; PF02787; CPSase_L_D3; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PRO0098; CPSASE.
 DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
 DR PROSITE; PS00866; CPSASE_1; 2.
 DR PROSITE; PS00867; CPSASE_2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; 3D-structure; Complete proteome.
 FT INIT MET 0
 FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 403 552 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 936 1072 ALLOSTERIC DOMAIN.
 FT REPEAT 1 552
 FT REPEAT 553 1072
 FT NP_BIND 152 209
 FT NP_BIND 302 353
 FT METAL 284 284
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 828 828
 FT METAL 840 840
 FT STRAND 8 12
 FT STRAND 18 18
 FT TURN 19 20
 FT STRAND 21 21
 FT TURN 23 23
 FT HELIX 24 38
 FT TURN 39 40
 FT STRAND 42 46
 FT TURN 50 51
 FT HELIX 53 55
 FT HELIX 57 59
 FT STRAND 62 64
 FT HELIX 70 80
 FT STRAND 84 86
 FT HELIX 91 103
 FT TURN 104 105
 FT HELIX 106 109
 FT TURN 110 111
 FT STRAND 113 114
 FT HELIX 119 126
 FT HELIX 128 137
 FT TURN 138 139
 FT STRAND 145 148
 FT HELIX 151 161
 FT STRAND 165 169
 FT TURN 170 171
 FT TURN 174 177
 FT STRAND 179 181
 FT HELIX 184 197
 FT TURN 199 200
 FT STRAND 203 207
 FT TURN 210 211
 FT STRAND 213 221
 FT TURN 223 224
 FT STRAND 227 235
 FT TURN 239 240

Query Match 19.0%; Score 61.5; DB 1; Length 1072;
 Best Local Similarity 29.0%; Pred. No. 53;
 Matches 18; Conservative 18; Mismatches 19; Indels 7; Gaps 2;
 QY 2 IDIIGTSPTSWEQAAAEAVQARDSDVDIRV-----ARVIEDQMAVDASGKITRYKLEVSF 56
 Db 658 VPIGTSPTDIR--AEDREFQHAVERLKLQKPNATVTAIEMAVEKAKEIGPLVLRP 715

QY 57 SF 58
 Db 716 SY 717

RESULT 9

RS7_BRAOL
 ID RS7_BRAOL STANDARD; PRT; 191 AA.
 AC Q9XK45;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 40S ribosomal protein S7.
 GN RPS7.
 OS Brassica oleracea (Cauliflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3712;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gossen K.K., Katavic V., Taylor D.C.;
 RT "Nucleotide sequence of a cDNA from microspore derived embryos of
 RT Brassica oleracea is homologous to 40S ribosomal protein S7.";
 RL (in) Plant Gene Register PGR99-096.
 CC -1- SIMILARITY: BELONGS TO THE S7E FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF144752; AAD44761.1; -
 DR InterPro; IPR000554; Ribosomal_S7E.
 DR Pfam; PF01251; Ribosomal_S7E; 1.
 DR ProDom; PD006276; Ribosomal_S7E; 1.
 DR PROSITE; PS00948; RIBOSOMAL_S7E; FALSE_NEG.
 KW Ribosomal protein.
 SQ SEQUENCE 191 AA; 22154 MW; 175A2D794210077C CRC64;

Query Match 18.8%; Score 61; DB 1; Length 191;
 Best Local Similarity 23.1%; Pred. No. 11;
 Matches 15; Conservative 15; Mismatches 23; Indels 12; Gaps 2;
 QY 6 GTSPTSWEQAAAEAV-----QARDSDVDIRVARDVQMAVDASGKITRYKLEVSF 58
 Db 13 GVAPTEFEERVAQAPDLENTQELKSLKLYINQAVSMDIAGNEKAVIY----VPP 67
 QY 59 KMRPA 63
 Db 68 RLAKA 72

RESULT 10
 IDI2_RHILO STANDARD; PRT; 351 AA.
 AC Q89L5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE [isopentenyl pyrophosphate isomerase].
 GN FNI OR M86371.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain)
 GN CARB OR STM0067.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RA Lu C.D., Walthall D.A., Abdelal A.T.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtneay L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.";
 RT Nature 413:852-856 (2001).
 RN [3]
 RP SEQUENCE OF 1-7 FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=88329100; PubMed=2843375;
 RA Kilstrup M., Lu C.D., Abdelal A., Neuhaard J.;
 RT "Nucleotide sequence of the *carA* gene and regulation of the *carAB* operon in *Salmonella typhimurium*.";
 RL Eur. J. Biochem. 176:421-429 (1988).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- PATHWAY: Pyrimidine biosynthesis.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate.
 CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; U81260; AAB39256.1; -.
 DR EMBL; AE008696; AAL19031.1; -.
 DR EMBL; M36540; AAB27033.1; -.
 DR HSSP; P00968; IAXX.
 DR StyGene; SGI0033; carB.
 DR HAMAP; MF 01210; -; 1.
 DR InterPro; IPR006275; CarA_L_glu.
 DR InterPro; IPR005483; CPase_L.
 DR InterPro; IPR005479; CPase_L_D2.
 DR InterPro; IPR005480; CPase_L_D3.
 DR InterPro; IPR005481; CPase_L_N.
 DR InterPro; IPR004362; MGS_like.
 DR Pfam; PF00289; CPase_L_chain; 2.
 DR Pfam; PF02786; CPase_L_D2; 2.
 DR Pfam; PF02787; CPase_L_D2; 2.
 DR Pfam; PF02142; MGS; 1.
 DR Pfam; PF02142; MGS; 1.
 DR PRINTS; PR00098; CPASASE.
 DR TIGRFAMs; TIGR01369; CPASASEII_lrg; 1.
 DR PROSITE; PS00866; CPASE_1; 2.
 DR PROSITE; PS00867; CPASE_2; 2.
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW

KW ATP-binding; Manganese; Complete proteome.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 403 552 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 936 1074 ALLOSTERIC DOMAIN.
 FT REPEAT 1 552
 FT REPEAT 553 1074 ATP (POTENTIAL).
 FT NP_BIND 152 209 ATP (POTENTIAL).
 FT NP_BIND 302 353 MANGANESE 1 (BY SIMILARITY).
 FT METAL 284 284 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
 FT METAL 300 300 MANGANESE 3 (BY SIMILARITY).
 FT METAL 828 828 MANGANESE 3 (BY SIMILARITY).
 FT METAL 840 840 MANGANESE 3 (BY SIMILARITY).
 FT CONFLICT 122 122 I -> G (IN REF. 1).
 SQ SEQUENCE 1074 AA; 118007 MW; 19243FECE2C8D5C7 CRC64;
 Query Match 18.7%; Score 60.5; DB 1; Length 1074;
 Best Local Similarity 29.0%; Pred. No. 68;
 Matches 18; Conservative 17; Mismatches 20; Indels 7; Gaps 2;
 QY 2 IDIIIGTSWQAARAEVQRADSVDDIRV-----ARVIEQDMAVDSAGKTYIRKLEV 56
 DB 658 VPIGTSPIADIR--AEDRERFCHAVDRILKQPNATVTAIEQAVEKAKEIGYPLVVRP 715
 QY 57 SF 58
 DB 716 SY 717
 RESULT 13
 K6PF PASMU STANDARD; PRT; 321 AA.
 ID K6PF PASMU STANDARD; PRT; 321 AA.
 AC QSCP2; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
 GN PHOSPHOHEXOKINASE.
 GN PFKA OR PM0069.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70.
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of *Pasteurella multocida* Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.
 CC -!- PATHWAY: Key control step of glycolysis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the phosphofructokinase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; AE006042; AAK02153.1; -.
 DR HSSP; P06998; 2PFK.
 DR HAMAP; MF 00339; -; 1.
 DR InterPro; IPR000023; PpfFructKinase.
 DR Pfam; PF00365; PFK; 1.
 DR PRINTS; PR00476; PHFRCTKINASE.
 DR ProDom; PD000707; PpfFructKinase; 1.

[illegible]

Db 833 VLSLGGADTLVTVVAAEWKYRAVEAVRRRERGASMKKEALREAFKVEGVDKREARLVQ 892

QY 38 Q 38

Db 893 Q 893

Search completed: November 21, 2003, 16:04:17
Job time : 1.77823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 , Search time 3.24762 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-6
Perfect score: 324
Sequence: 1 VIDIIIGTSPISWEQAAAEAV.....KITVRIKLEVSFKWRPAQPR 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriaph:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	71	16	053714
2	319	98.5	116	2	007267
3	265	81.8	65	2	007265
4	190	58.6	110	2	006124
5	140	43.2	96	2	006125
6	101	31.2	77	17	Q9HPW4
7	92	28.4	110	16	Q8RAA6
8	76	23.5	70	16	Q8KAD4
9	69	21.3	71	16	Q9I794
10	66	20.4	69	16	Q92X94
11	65	20.1	83	16	Q8XQ96
12	64.5	19.9	501	6	Q8WNL1
13	64.5	19.9	1030	5	Q362B3
14	64	19.8	96	16	Q8Z0H1
15	64	19.8	206	15	Q75741
16	64	19.8	523	5	Q9NE33

17	64	19.8	1180	3	Q8J0Z4
18	63	19.8	1200	3	Q8J122
19	63	19.4	85	16	Q9RCZ5
20	63	19.4	155	12	Q39829
21	63	19.4	526	3	Q8X0K0
22	62.5	19.3	436	2	Q8RPT8
23	62.5	19.3	456	12	Q8JL22
24	62	19.1	282	2	O52068
25	62	19.1	282	16	Q8YDMS
26	62	19.1	282	16	Q8FUS2
27	62	19.1	1090	2	Q8RPF1
28	62	19.1	1094	5	Q23915
29	61.5	19.0	252	16	Q32LV3
30	61.5	19.0	284	2	Q9AE26
31	61.5	19.0	274	12	Q84181
32	61.5	19.0	335	5	Q8T3W1
33	61.5	19.0	339	5	Q8INJ1
34	61.5	19.0	477	5	Q8INJ2
35	61.5	19.0	482	16	Q8FQL4
36	61.5	19.0	495	5	Q9VGD7
37	61.5	19.0	1073	16	Q8FLB0
38	61.5	19.0	1477	5	Q9VB52
39	61.5	19.0	1477	5	O76931
40	61	18.8	94	17	Q9HRH0
41	61	18.8	397	2	Q9AG69
42	61	18.8	2626	11	Q32LN3
43	60.5	18.7	373	2	Q8KJK8
44	60.5	18.7	481	2	Q9X754
45	60.5	18.7	662	9	Q9G064

ALIGNMENTS

RESULT 1

OS3714 ID OS3714 PRELIMINARY; PRT; 71 AA.
AC OS3714;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Possible protein TRANSPORT protein.
GN SEC OR RV0379 OR MT036.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544 (1998).
RL EMBL; AL021931; CAA17385.1; --
DR TuberculList; RV0379; --
KW Complete proteome.
SQ SEQUENCE 71 AA; 7966 MW; 49526FDBCA9C3826 CRC64;

Query Match 100.0%; Score 324; DB 16; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VIDIIIGTSPISWEQAAAEAVQSRQSDVDIRVARVIEDQMAVDASAGKITVRIKLEVSFKWRPAQPR 60
|||||

```
Db 6 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 65
QY 61 RPAQPR 66
Db 66 RPAQPR 71

RESULT 2
O07267 PRELIMINARY; PRT; 116 AA.
AC O07267;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 12.7 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TB patient sputum #NTI64719;
RA Rao A.R., Vijaya S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85467; AAB58554.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 116 AA; 12673 MW; DF56D2FF0B479896 CRC64;

Query Match 98.5%; Score 319; DB 2; Length 116;
Best Local Similarity 98.5%; Pred. No. 1e-28;
Matches 65; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
Db 51 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 110
QY 61 RPAQPR 66
Db 111 RPAQPR 116

RESULT 3
O07265 PRELIMINARY; PRT; 65 AA.
AC O07265;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 7.2 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TB patient sputum #NTI64719;
RA Rao A.R., Vijaya S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85465; AAB58552.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 65 AA; 7240 MW; 9939C19E2306FA1B CRC64;

Query Match 81.8%; Score 265; DB 2; Length 65;
Best Local Similarity 85.9%; Pred. No. 7.3e-23;
Matches 55; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 IDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 61
Db 1 IDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYIKLEVSFKM 60
QY 62 PAQP 65
```

```
Db 61 PAQP 64

RESULT 4
O06124 PRELIMINARY; PRT; 110 AA.
AC O06124;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 11.9 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTI 64719;
RA Sachidanandam V., Rama Rao A., Hegde S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001301; AAB54027.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 110
SQ SEQUENCE 110 AA; 11903 MW; A21AB7D99A7B4F54 CRC64;

Query Match 58.6%; Score 190; DB 2; Length 110;
Best Local Similarity 95.2%; Pred. No. 4.8e-14;
Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAV 42
Db 53 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARVIEQDMAV 94

RESULT 5
O06125 PRELIMINARY; PRT; 96 AA.
AC O06125;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE Hypothetical 10.3 kDa protein (Fragment).
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTI 64719;
RA Sachidanandam V., Rama Rao A., Hegde S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001302; AAB54028.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10298 MW; 2E5B14CB1D1E445C CRC64;

Query Match 43.2%; Score 140; DB 2; Length 96;
Best Local Similarity 86.1%; Pred. No. 2e-08;
Matches 31; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARV 36
Db 61 VIDIIGTPTSWEQAAAEAVQARSDVDDIRVARV 96

RESULT 6
Q9HPW4 PRELIMINARY; PRT; 77 AA.
ID Q9HPW4
AC Q9HPW4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
```

01-MAR-2001 (TRENBLrel. 16, Last sequence update)
01-JUN-2001 (TRENBLrel. 17, Last annotation update)
Vng1446h.
VNG1446H.
GN Halobacterium sp. (strain NRC-1).
OC Archaea: Euryarchaeota: Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welci R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Studich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RA "Genome sequence of Halobacterium species NRC-1".
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
DR ENBL; AB005061; AG19753.1; -.
KW Complete proteome.
SQ SEQUENCE 77 AA; 8378 MW; C3C359B65E8A8306 CRC64;
Query Match 31.2%; Score 101; DB 17; Length 77;
Best Local Similarity 31.1%; Pred. No. 0.00044;
Matches 19; Conservative 22; Mismatches 20; Indels 0; Gaps 0;
QY 4 IIGTSTSEQAAAEAVQARSVDIRVARVIEDMAVDSAGKITRIKLEVSFKM 63
Db 17 LGTSEESPTAAADDAIDRAEDTLNVVMAEVVDQGVIGAEVETRTQTVQVAFELDGS 76
QY 64 Q 64
Db 77 Q 77
RESULT 7
Q8RAA6 PRELIMINARY; PRT; 110 AA.
ID Q8RAA6
AC Q8RAA6;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein TTE1319.
GN TTE1318.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome".
RL Genome Res. 12:689-700 (2002).
DR ENBL; AB013092; AAM24542.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 110 AA; 11862 MW; 43C0606A4B05F6DB CRC64;
Query Match 28.4%; Score 92; DB 16; Length 110;
Best Local Similarity 28.8%; Pred. No. 0.0071;
Matches 17; Conservative 17; Mismatches 25; Indels 0; Gaps 0;
QY 1 VIDIGTSTSEQAAAEAVQARSVDIRVARVIEDMAVDSAGKITRIKLEVSFK 59
Db 46 VLVNVDGDSVTSVEDAIHRAVEAAKSVNIGSIEVNVGTANVKGKIVEKANIQIAYR 104
RESULT 8
Q8KAD4 PRELIMINARY; PRT; 70 AA.
ID Q8KAD4
AC Q8KAD4;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein CT2229.
GN CT2229.
OS Chlorobium tepidum.
OC Bacteria: Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobiaceae; Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Bisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.P., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feidblyum T.V., Hansen C.L., Gruber T.M., Ketchum K.A.,
RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic green-sulfur bacterium".
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
DR ENBL; AE012968; AAM73445.1; -.
DR TIGR; CT2229; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 7922 MW; 8C662B18D044F153 CRC64;
Query Match 23.5%; Score 76; DB 16; Length 70;
Best Local Similarity 30.0%; Pred. No. 0.028;
Matches 18; Conservative 19; Mismatches 21; Indels 2; Gaps 2;
QY 2 IDIIGTSTSEQAAAEAVQARSVDIRVARVIEDMAVDSAGKITRIKLEVSFKM 60
Db 9 LEIVGSSATSTIEEAVNNVAKAETIRNIRVELVETRCHVENQ-KIAYVQVTKIGFTL 67
RESULT 9
Q91794 PRELIMINARY; PRT; 71 AA.
ID Q91794
AC Q91794;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE Hypothetical protein PA0038.
GN PA0038.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen".
RL Nature 406:959-964 (2000).
DR ENBL; AE004443; AAG03428.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 71 AA; 7863 MW; 5925D78BE2709A5E CRC64;
Query Match 21.3%; Score 69; DB 16; Length 71;
Best Local Similarity 25.4%; Pred. No. 1.8;

```
Matches 15; Conservative 17; Mismatches 27; Indels 0; Gaps 0;
QY 2 IDIIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 60
Db 10 IELVSSKTSIEDANNALEAAKSIQHLFEVVDTRGHIENGAVGHYQVILKVGFR 68

RESULT 10
Q92X94 PRELIMINARY; PRT; 69 AA.
AC Q92X94;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein RB0068.
GN RB0068 OR SMB20068.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
DR EMBL; AL603642; CAC48468.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 69 AA; 7792 MW; F53BFB80AE62C07D CRC64;

Query Match 20.4%; Score 66; DB 16; Length 69;
Best Local Similarity 27.9%; Pred. No. 3.8;
Matches 19; Conservative 17; Mismatches 14; Indels 18; Gaps 4;
QY 2 IDIIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRI 52
Db 9 IELIGSPNSIDEAIEGASIRASKITRNLDFEVDQIR-GQIVN-----GKVARYQV 59

QY 53 KLEVSFPM 60
Db 60 VMKVGPR 67

RESULT 11
Q8XQP6 PRELIMINARY; PRT; 83 AA.
AC Q8XQP6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Rsp1175.
GN Rsp1175 OR RS05065.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Ariguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
```

```
RL Nature 415:497-502 (2002).
DR EMBL; AL646083; CAD18326.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 83 AA; 9140 MW; 1EAAD51D4FA4871A CRC64;

Query Match 20.1%; Score 65; DB 16; Length 83;
Best Local Similarity 28.8%; Pred. No. 6.1;
Matches 19; Conservative 20; Mismatches 23; Indels 4; Gaps 3;
QY 1 VIDIGTSPTEQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 58
Db 8 MIELVSSPESDAAIENAIARASETLHYLDWFVWETRGHLVD--GKIAHQVILKVG 65

QY 59 KMRPAQ 64
Db 66 RLQEAQ 71

RESULT 12
Q8WVN1 PRELIMINARY; PRT; 501 AA.
AC Q8WVN1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glucose transporter 5.
GN SUC2A5.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSU=Small intestine;
RA Wood I.S., Castano-Mereditz E.F., Dyer J., Shirazi-Beechey S.P.;
RT "Nutrient regulation of GLUT5 in sheep intestine.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AJ315928; CAC86964.1; -.
DR InterPro; IPR000005; HTHArac.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub transporter.
DR InterPro; IPR003663; Sugar transp.
DR InterPro; IPR005829; Sug transporter.
DR Pfam; PF00083; sugar tr; 1.
DR PRINTS; PR00171; SUGTRNSPORT.
DR TIGRFAMs; TIGR00879; SP; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS00850; MFS; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Sugar transport; Transmembrane; Transport.
SQ SEQUENCE 501 AA; 55546 MW; 23F0E43E944C9AD5 CRC64;

Query Match 19.9%; Score 64.5; DB 6; Length 501;
Best Local Similarity 40.0%; Pred. No. 56;
Matches 20; Conservative 10; Mismatches 15; Indels 5; Gaps 3;
QY 13 EQAAAEVQARDSDVDIRVARVIEQDMAVDSAGKIT-YRIKLEVSF 61
Db 227 EEAARALRLRGHWHDVDAIEIIEILBEDRAEKAAGFISV-LKL---FKMR 272

RESULT 13
Q962B3 PRELIMINARY; PRT; 1030 AA.
AC Q962B3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aminopeptidase N (EC 3.4.11.2).
```

```
GN APN2.
OS Helicoverpa armigera (Cotton bollworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Heliothinae; Helicoverpa.
OX NCBI_TaxID=29059;
RN [1]
RN SEQUENCE FROM N.A.
RP Angelucci C., Akhurst R.J., East P.;
RA "Helicoverpa armigera aminopeptidase N gene APN2.";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY038608; AAK85539.1; -.
DR MEROPS; M01.013; -.
DR InterPro; IPR001930; Ala_peptase.
DR InterPro; IPR006025; Zn_MPeptidase.
DR Pfam; PF01433; Peptidase M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Aminopeptidase; Hydrolase.
SQ SEQUENCE 1030 AA; 114505 MW; 258A43E74EFBFA7 CRC64;

Query Match 19.8%; Score 64.5; DB 5; Length 1030;
Best Local Similarity 30.4%; Pred. No. 1.3e+02;
Matches 14; Conservative 14; Mismatches 17; Indels 1; Gaps 1;

QY 11 SWEQAAAEAVQARSDVDDIRVARVIEQDVAVDVDSAGKITRYIKLEY 56
Db 610 NWELLAA-ALSRDHTAIHLNRAQIVDDVFALMRSGQITRYLGFKV 654

RESULT 14
Q8Z0H1 PRELIMINARY; PRT; 96 AA.
AC Q8Z0H1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional regulator.
GN ASL0125.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shampo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
PL DNA Res. 8:205-213(2001).
DR EMBL; AP003581; BAB77649.1; -.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SMO0530; HTH_XRE; 1.
KW Complete proteome.
SQ SEQUENCE 96 AA; 10897 MW; AC99D93E3B06C901 CRC64;

Query Match 19.8%; Score 64; DB 16; Length 96;
Best Local Similarity 26.7%; Pred. No. 9.4;
Matches 16; Conservative 10; Mismatches 28; Indels 6; Gaps 1;

QY 7 TSPTSWEQAAAEAVQARSDVDDIRVARVIEQDVAVDVDSAGKITRYIKLEYSFKMRPAQPR 66
Db 26 TSVRWAEGRTEPTWPKKYVDALRI PKCTPEELASASEKSINQR-----HKKPGRRP 79

RESULT 15
Q75741 PRELIMINARY; PRT; 206 AA.
ID Q75741
AC Q75741;
```

```
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nef (Negative factor) (P-protein) (27 kDa protein)
DE (Fragment).
DE NEF.
GN Human immunodeficiency virus 1.
OS Viruses; Retrod viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=HP93cl;
RA Mariani R., Kirchhoff F., Greenough T.C., Sullivan J.L.,
RA Desrosiers R.C., Skowronski J.;
RT "High frequency of defective nef alleles in a long-term survivor with
nonprogressive HIV-1 infection.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
CC ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4(T4) ANTIGEN (BY
CC SIMILARITY).
DR EMBL; U61813; AAB04732.1; -.
DR HSSP; P03406; IEFN.
DR InterPro; IPR001558; HIV_Nef.
DR Pfam; PF00469; F-protein; 1.
DR ProDom; PD000031; HIV_Nef; 1.
KW AIDS; GTP-binding; Lipoprotein; Myristate.
FT NON_TER 1
SQ SEQUENCE 206 AA; 23974 MW; 12410737A86F8F25 CRC64;

Query Match 19.8%; Score 64; DB 15; Length 206;
Best Local Similarity 31.0%; Pred. No. 23;
Matches 18; Conservative 12; Mismatches 18; Indels 10; Gaps 2;

QY 18 EAVQARSDVDDIR-VARVIEQDVAVDVDSAGKITRYI-----KLVSEKMRPAQP 65
Db 18 ERMRAEPAADKVRVSRVLEKHGAI TSNTAANNVDCAWLEAQEEEEVSFPVRQVP 75

Search completed: November 21, 2003, 16:08:10
Job time : 5.24762 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 5.7517 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFERISGDL.....VOYSRADDEQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	AAW32444	Mycobacterium tube
2	466	100.0	95	AAW32376	Mycobacterium tube
3	466	100.0	95	AAW81747	M. tuberculosis im
4	466	100.0	95	AAW64321	Mycobacterium tube
5	466	100.0	95	AAW32097	Mycobacterium tube
6	466	100.0	95	AAW39118	M. tuberculosis an
7	466	100.0	95	AAW38981	M. tuberculosis re
8	466	100.0	95	AAE29717	Mycobacterium tube
9	466	100.0	95	AAE17581	Mycobacterium spec

10	466	100.0	100	19	AAW81706	M. tuberculosis im
11	466	100.0	100	19	AAW64339	Mycobacterium tube
12	466	100.0	100	20	AAW39136	M. tuberculosis an
13	466	100.0	100	20	AAW38993	M. tuberculosis re
14	466	100.0	100	22	AAW35218	M. tuberculosis RV3
15	466	100.0	100	22	AAW19845	Mycobacterium tube
16	466	100.0	100	23	ABU05988	M. tuberculosis an
17	466	100.0	100	23	ARG30969	Mycobacterium tube
18	466	100.0	358	23	AAU74591	Antigenic fusion p
19	466	100.0	802	19	AAW81746	M. tuberculosis fu
20	466	100.0	802	19	AAW64379	Mycobacterium anti
21	466	100.0	802	20	AAW32063	Mycobacterium tube
22	466	100.0	802	20	AAW39224	M. tuberculosis fu
23	466	100.0	802	20	AAW39176	M. tuberculosis fu
24	466	100.0	802	20	AAW39081	M. tuberculosis fu
25	466	100.0	802	20	AAW39033	M. tuberculosis fu
26	466	100.0	802	23	AAU74592	Antigenic fusion p
27	466	100.0	983	22	AAU01901	M. tuberculosis TB
28	462	99.1	100	20	AAW03705	M. tuberculosis LH
29	392	84.1	80	18	AAW32454	Mycobacterium tube
30	392	84.1	80	18	AAW32386	Mycobacterium tube
31	392	84.1	80	19	AAW81707	M. tuberculosis im
32	392	84.1	80	19	AAW64340	Mycobacterium tube
33	392	84.1	80	20	AAW39137	M. tuberculosis an
34	392	84.1	80	20	AAW38994	M. tuberculosis re
35	217	46.6	49	20	AAW03706	M. tuberculosis LH
36	205	44.0	42	20	AAW03707	M. tuberculosis LH
37	173	37.1	100	23	ABU05361	M. tuberculosis an
38	137	29.4	28	20	AAW03712	M. tuberculosis LH
39	122	26.2	28	18	AAW32460	Mycobacterium tube
40	122	26.2	28	19	AAW81698	M. tuberculosis im
41	122	26.2	28	20	AAW39128	M. tuberculosis an
42	121	26.0	27	18	AAW32458	Mycobacterium tube
43	121	26.0	27	19	AAW81696	M. tuberculosis im
44	121	26.0	27	20	AAW39126	M. tuberculosis an
45	118	25.3	27	18	AAW32457	Mycobacterium tube

ALIGNMENTS

RESULT 1
AAW32444
ID AAW32444 standard; Protein; 95 AA.
XX
AC AAW32444;
DT 09-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TB38-1.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WC9709428-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14674.
XX
PR 12-JUL-1996; 96US-0680574.
PR 01-SEP-1995; 95US-0523436.
PR 22-SEP-1995; 95US-0533634.
PR 22-MAR-1996; 96US-0620874.
PR 05-JUN-1996; 96US-0659683.
(CORI-) CORIXA CORP.
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
XX

DR WPI; 1997-192903/17.
 XX N-PSDB; AAT91509.
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 XX Example 3; Page 124; 168pp; English.
 XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX Sequence 95 AA;
 SQ Query Match 100.0%; Score 466; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 DB 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 QY 61 KOELDEISTNIRAGVQYSRADDEEQOALSSQMGF 95
 DB 61 KOELDEISTNIRAGVQYSRADDEEQOALSSQMGF 95
 RESULT 2
 AAW32376
 ID AAW32376 standard; Protein; 95 AA.
 XX AAW32376;
 AC AAW32376;
 DT 13-JAN-1998 (first entry)
 XX Mycobacterium tuberculosis antigen Tb38-1.
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 OS Mycobacterium tuberculosis.
 XX WO9709429-A2.
 PN 13-MAR-1997.
 PD 30-AUG-1996; 96WO-US14675.
 XX 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX (CORI-) CORIXA CORP.
 PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX WPI; 1997-192904/17.
 DR N-PSDB; AAT91445.
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX Sequence 95 AA;
 SQ Query Match 100.0%; Score 466; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 DB 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFEAAANKQ 60
 QY 61 KOELDEISTNIRAGVQYSRADDEEQOALSSQMGF 95
 DB 61 KOELDEISTNIRAGVQYSRADDEEQOALSSQMGF 95
 RESULT 3
 AAW81747
 ID AAW81747 standard; Protein; 95 AA.
 XX AAW81747;
 AC AAW81747;
 DT 27-JAN-1999 (first entry)
 XX M. tuberculosis immunogenic polypeptide Tb38-1.
 DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.
 OS Mycobacterium tuberculosis.
 XX WO9816646-A2.
 PN 23-APR-1998.
 PD 07-OCT-1997; 97WO-US18293.
 XX 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX (CORI-) CORIXA CORP.
 PA Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64491.
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX Example 3b; Page 117; 230pp; English.
 PS This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 XX Sequence 95 AA;


```
Query Match      100.0%; Score 466; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95

RESULT 4
AAW64321
ID AAW64321 standard; Peptide; 95 AA.
AC AAW64321;
DT 09-NOV-1998 (first entry)
DE Mycobacterium tuberculosis antigen Tb38-1 peptide.
XX Tuberculosis; infection; diagnosis; antigen; Tb38-1.
KW Mycobacterium tuberculosis strain H37Rv.
OS WO9816645-A2.
FN 23-APR-1998.
PD 07-OCT-1997; 97WO-US18214.
PF 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX (CORI-) CORIXA CORP.
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR N-PSDB; AAW44384.
XX WPI; 1998-251292/22.
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
PS Example 3; Page 123; 250pp; English.
XX This is an antigenic portion of Mycobacterium tuberculosis antigen
CC Tb38-1. A DNA sequence (see AAW44384) coding for antigen Tb38-1 was
CC isolated from a M. tuberculosis strain H37Rv expression library
CC using sera from patients having pulmonary or pleural tuberculosis.
CC The invention relates to compositions and methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAW64291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
CC as well as DNA sequences encoding such polypeptides, recombinant
CC expression vectors and transformed or transfected host cells. Also
CC claimed are methods and diagnostic kits for detecting M.
CC tuberculosis infection in a patient using these polypeptides,
CC antibodies or oligonucleotide probes and primers, for the diagnosis
CC of tuberculosis.
XX
SQ Sequence 95 AA;
Query Match      100.0%; Score 466; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95

RESULT 6
AAW39118
ID AAW39118 standard; Peptide; 95 AA.
XX
```

```
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95

RESULT 5
AAW32097
ID AAW32097 standard; Protein; 95 AA.
XX AAW32097;
XX 17-JAN-2000 (first entry)
DT Mycobacterium tuberculosis antigen Tb38-1.
DE Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;
XX vaccine; immunogen.
XX Mycobacterium tuberculosis.
XX WO9951748-A2.
FN 14-OCT-1999.
PD 07-APR-1999; 99WO-US07717.
PF 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
XX New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
PS Claim 1; Fig 4D; 83pp; English.
XX This sequence represents the Mycobacterium tuberculosis antigen
CC Tb38-1. The invention provides fusion proteins (see AAY32059-71)
CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.
CC TBP-2 (see AAY32063) and a TbH9-Tb38-1 fusion. The new fusion
CC proteins are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.
XX
SQ Sequence 95 AA;
Query Match      100.0%; Score 466; DB 20; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95

RESULT 6
AAW39118
ID AAW39118 standard; Protein; 95 AA.
XX
```

AC AAY39118;
 XX 05-NOV-1999 (first entry)
 DT
 XX M. tuberculosis antigen Tb38-1 amino acid sequence.
 DE
 XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 PN
 PD 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-US03268.
 PF
 XX 05-MAY-1998; 98US-0072967.
 PR
 XX 18-FEB-1998; 98US-0025197.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 PI
 XX WPI; 1999-527409/44.
 DR
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 PT
 XX Example 3; Page 113; 299pp; English.
 PS
 XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T. B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 CC
 XX Sequence 95 AA;
 SQ

Query Match 100.0%; Score 466; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEANKQ 60
 DB 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEANKQ 60
 QY 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 7
 AAY38981
 ID AAY38981 standard; Protein; 95 AA.
 XX
 AC AAY38981;
 XX
 XX 05-NOV-1999 (first entry)
 DT
 XX M. tuberculosis recombinant antigen protein Tb38-1.
 DE
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX

OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 PN
 PD 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-US03265.
 PF
 XX 05-MAY-1998; 98US-0072596.
 PR
 XX 18-FEB-1998; 98US-0024753.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 PI
 XX WPI; 1999-527416/44.
 DR
 XX N-FSDB; AAZ19082.
 DR
 XX New polypeptide comprising antigenic portions of M. tuberculosis
 PT
 XX Example 3; Page 159; 323pp; English.
 PS
 XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 CC
 XX Sequence 95 AA;
 SQ

Query Match 100.0%; Score 466; DB 20; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.1e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEANKQ 60
 DB 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEANKQ 60
 QY 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
 DB 61 KQELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 8
 AAE29717
 ID AAE29717 standard; Protein; 95 AA.
 XX
 AC AAE29717;
 XX
 XX 27-JAN-2003 (first entry)
 DT
 XX Mycobacterium tuberculosis Tb38-1 antigenic protein.
 DE
 XX Vaccine; immunity; diagnostic agent; gene therapy; Tb38-1 antigen;
 KW MTb11.
 KW
 XX Mycobacterium tuberculosis.
 OS
 XX WO200272792-A2.
 PN
 XX 19-SEP-2002.
 PD
 XX 13-MAR-2002; 2002WO-US08223.
 PF
 XX 13-MAR-2001; 2001US-275837P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Brannon M, Guderian J;
 PI
 XX

DR WPI; 2002-759844/82.
DR N-PSDB; AAD47094.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA,
PT LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
PT immunity against pathogenic microorganisms e.g. Leishmania and
PT Mycobacterium tuberculosis
XX
PS Disclosure; Page 107; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from T8A, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides,
CC as in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC M. tuberculosis Tb38-1 antigenic protein. Tb38-1 is also referred to
CC as MTb11 or 38-1.
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 466; DB 23; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
RESULT 9
AAE17581
ID AAE17581 standard; Protein; 95 AA.
AC AAE17581;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species Tb38-1 (Mtbi1; 38-1) protein.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; Tb38-1; Mtbi1; 38-1 protein.
XX
OS Mycobacterium sp.
XX
FN WO2001198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US19959.
XX
PR 20-JUN-2000; 2000US-0597796.
PR 01-FEB-2001; 2001US-265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
DR N-PSDB; AAD28352.
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a

PT subject -
XX
PS Claim 9; Page 123; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC Tb38-1 (Mtbi1; 38-1) protein.
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 466; DB 23; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQ 60
QY 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQVSRADDEEQQALSSQMGF 95
RESULT 10
AAW81706
ID AAW81706 standard; Protein; 100 AA.
XX
AC AAW81706;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb38-IN.
XX
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
FN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1998-261042/23.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Example 3B; Page 138-139; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This
 CC sequence can be formulated into vaccines and/or pharmaceutical
 CC compositions for immunising against M. tuberculosis infection or may
 CC be used for the diagnosis of tuberculosis.

XX Sequence 100 AA;

Query Match 100.0%; Score 466; DB 19; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 DB 6 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 DB 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 11

AAW64339
 ID AAW64339 standard; Protein; 100 AA.

XX AC AAW64339;

DT 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-IN.

XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.

XX Mycobacterium tuberculosis strain H37RV.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis

XX Example 3; Page 145; 250pp; English.

XX This polypeptide comprises a partial sequence of Mycobacterium
 CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated
 CC from a M. tuberculosis strain H37RV genomic library using a probe
 CC derived from clone Tb38-1 (see AAW44384). The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of
 CC a soluble M. tuberculosis antigen, or an immunogenic portion of an
 CC M. tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic

CC kits for detecting M. tuberculosis infection in a patient using
 CC these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis.

XX Sequence 100 AA;

Query Match 100.0%; Score 466; DB 19; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 DB 6 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95

DB 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 12

AAAY39136
 ID AAY39136 standard; Protein; 100 AA.

XX AC AAY39136;

DT 05-NOV-1999 (first entry)

XX M. tuberculosis antigen Tb38-IN amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

OS Mycobacterium tuberculosis.

XX WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025137.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions

XX Example 3; Page 133-134; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 100 AA;

Query Match 100.0%; Score 466; DB 20; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65
QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 13
AAV38993
ID AAV38993 standard; Protein; 100 AA.
XX AC
XX AAV38993;
XX 05-NOV-1999 (first entry)
XX DE M. tuberculosis recombinant antigen protein Tb38-IN.
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
XX KW vaccine; immunity.
XX OS Mycobacterium tuberculosis.
XX PN WO9942118-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 98WO-US03265.
XX PR 05-MAY-1998; 98US-0072596.
XX PR 18-FEB-1998; 98US-0024753.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527416/44.
XX PT New polypeptide comprising antigenic portions of M. tuberculosis
XX PS Example 3; Page 179; 323pp; English.
XX CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX SQ Sequence 100 AA;

Query Match 100.0%; Score 466; DB 20; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65
QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 14
AAB35218
ID AAB35218 standard; Protein; 100 AA.
XX AC
XX AAB35218;

```

```

XX 24-APR-2001 (first entry)
XX DE M tuberculosis Rv3874 protein.
XX KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c;
XX KW Rv1037c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c;
XX KW Rv3445c; Rv3890c; Rv3891c; Rv3904c; Rv3905c.
XX OS Mycobacterium tuberculosis.
XX PN WO200104151-A2.
XX PD 18-JAN-2001.
XX PF 13-JUL-2000; 2000WO-DK00398.
XX PR 13-JUL-1999; 99DK-0001020.
XX PR 15-JUL-1999; 99US-0144011.
XX PA (STAT-) STATENS SERUM INST.
XX PI Andersen P, Skjot R;
XX DR WPI; 2001-091923/10.
XX PT New polypeptide encoded by a member of the esat-6-gene family for
XX PS immunizing against and diagnosis of tuberculosis.
XX Example 2; Page 65; 80pp; English.
XX CC The present invention provides the protein and coding sequences for
XX CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
XX CC proteins include Rv0287, Rv1036c, Rv1037c, Rv2348c, Rv2653c,
XX CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and
XX CC Rv3905c. These can be used to produce vaccines against, and in the
XX CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
XX CC the proteins of the invention.
XX SQ Sequence 100 AA;

Query Match 100.0%; Score 466; DB 22; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLAQAGNFERISGLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65
QY 61 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 15
AAB19845
ID AAB19845 standard; Protein; 100 AA.
XX AC
XX AAB19845;
XX 05-MAR-2001 (first entry)
XX DE Mycobacterium tuberculosis protein MTEN4.
XX KW MTEN4; tuberculosis; BCG; vaccine; infection; diagnosis.
XX OS Mycobacterium tuberculosis.
XX PN WO200006157-A1.
XX PD 09-NOV-2000.
XX PF 04-MAY-2000; 2000WO-US12257.

```

```

XX 04-MAY-1999; 99US-0132505.
XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
XX Gennaro ML;
XX WPI: 2001-007153/01.
XX N-PSDB; AAA89038.
XX Novel polypeptide encoded by open reading frames present in
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
PT bovis, useful as vaccine and for diagnosing tuberculosis infection
XX Claim 11; Fig 1; 35pp; English.
XX The present sequence is that of the Mycobacterium tuberculosis
CC MTEN4 protein. This is 1 of 8 proteins, i.e. MTEN1-8 (see
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)
CC identified as being present in the genome of M. tuberculosis but
CC absent from the genome of the BCG strain of Mycobacterium bovis.
CC MTEN1-8 represent reagents that are useful in discriminating between
CC M. tuberculosis and BCG and, in particular, for diagnostic methods
CC which discriminate between exposure of a subject to M. tuberculosis
CC and vaccination with BCG. The invention features these MTEN
CC polypeptides, functional fragments of them, DNA encoding them,
CC vectors, transformed cells, and diagnostic, therapeutic, and
CC prophylactic (vaccine) methods, including genetic vaccination
XX methods.
XX SQ Sequence 100 AA;
Query Match 100.0%; Score 466; DB 22; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDAATLAQAGNFRISGDKTQIDOVETAGSLQGWEGAGCTAAQAARVRFQEAANKQ 60
Db 6 TDAATLAQAGNFRISGDKTQIDOVETAGSLQGWEGAGCTAAQAARVRFQEAANKQ 65
Qy 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
Db 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

```

Search completed: November 21, 2003, 16:03:12
Job time : 5.7517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 2.0034 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFERISGDL.....VOYSRADDEQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	3	US-08-818-112-88
2	466	100.0	95	4	US-08-818-111-89
3	466	100.0	95	4	US-09-056-556-88
4	466	100.0	95	4	US-09-072-596-89
5	466	100.0	100	3	US-08-818-112-115
6	466	100.0	100	4	US-08-818-111-110
7	466	100.0	100	4	US-09-056-556-115
8	466	100.0	100	4	US-09-072-596-110
9	466	100.0	802	4	US-09-056-556-214
10	466	100.0	802	4	US-09-072-596-209
11	466	100.0	802	4	US-09-072-596-346
12	462	99.1	100	4	US-09-116-492A-5
13	392	84.1	180	3	US-08-818-112-117
14	392	84.1	80	4	US-08-818-111-112
15	392	84.1	80	4	US-09-056-556-117
16	392	84.1	80	4	US-09-072-596-112
17	217	46.6	49	4	US-09-116-492A-6
18	205	44.0	42	4	US-09-116-492A-7
19	173	37.1	100	4	US-09-116-492A-28
20	137	29.4	28	4	US-09-116-492A-12
21	122	26.2	28	3	US-08-818-112-98
22	122	26.2	28	4	US-09-056-556-98
23	121	26.0	27	3	US-08-818-112-96
24	121	26.0	27	4	US-09-056-556-96
25	118	25.3	27	3	US-08-818-112-95
26	118	25.3	27	3	US-08-818-112-97
27	118	25.3	27	4	US-09-056-556-95

28	118	25.3	27	4	US-09-056-556-97	Sequence 97, Appl
29	117	25.1	28	3	US-08-818-112-93	Sequence 93, Appl
30	117	25.1	28	4	US-09-056-556-93	Sequence 93, Appl
31	100	21.5	20	4	US-09-116-492A-8	Sequence 8, Appl
32	87.5	18.8	97	4	US-09-073-009-109	Sequence 109, App
33	87.5	18.8	97	4	US-09-073-009-143	Sequence 143, App
34	85	18.2	16	3	US-08-818-112-94	Sequence 94, Appl
35	85	18.2	16	4	US-09-056-556-94	Sequence 94, Appl
36	81.5	17.5	955	1	US-08-006-676B-1	Sequence 1, Appl
37	81.5	17.5	955	1	US-08-282-845-2	Sequence 1, Appl
38	81.5	17.5	955	2	US-08-428-414A-3	Sequence 3, Appl
39	81.5	17.5	955	5	PCT-US94-00324-1	Sequence 1, Appl
40	79	17.0	16	4	US-09-116-492A-10	Sequence 10, Appl
41	78	16.7	16	4	US-09-116-492A-9	Sequence 9, Appl
42	78	16.7	16	4	US-09-116-492A-13	Sequence 13, Appl
43	77	16.5	16	4	US-09-116-492A-11	Sequence 11, Appl
44	77	16.5	336	4	US-09-252-991A-24656	Sequence 24656, A
45	75	16.1	1147	1	US-08-144-121-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-08-818-112-88
; Sequence 88, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-88

Query Match 100.0%; Score 466; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 2
US-08-818-111-89
; Sequence 89, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/818,111
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-89

Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2,1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 3
US-09-056-556-88
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/056,556
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-88

Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2,1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
Db 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAVVRFQEAANKQ 60
QY 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95
Db 61 KQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 95

RESULT 4
US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/072.596
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-072-596-89
Query Match 100.0%; Score 466; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.1e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95

RESULT 5
US-08-818-112-115
; Sequence 115, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-818-111-110
Query Match 100.0%; Score 466; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95

RESULT 6
US-08-818-111-110
; Sequence 110, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-818-111-110
Query Match 100.0%; Score 466; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60
Db 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60

QY 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100

US-08-818-112-115
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-818-112-115
Query Match 100.0%; Score 466; DB 3; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.3e-46;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 60
Db 6 TDAATLAQAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAAKQ 65

QY 61 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95
Db 66 KOELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100
```

```
QY 61 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 100

RESULT 7
US-09-556-556-115
; Sequence 115, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; QUERY MATCH
; BEST LOCAL SIMILARITY 100.0%; Score 466; DB 4; Length 100;
; Mismatches 0; Indels 0; Gaps 0;
; Query Match 100.0%; Score 466; DB 4; Length 100;
; Best Local Similarity 100.0%; Pred. No. 2.3e-46;
; Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 60
DB 6 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 65
QY 61 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 100

RESULT 8
US-09-072-596-110
; Sequence 110, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; QUERY MATCH
; BEST LOCAL SIMILARITY 100.0%; Score 466; DB 4; Length 100;
; Mismatches 0; Indels 0; Gaps 0;
; Query Match 100.0%; Score 466; DB 4; Length 100;
; Best Local Similarity 100.0%; Pred. No. 2.3e-46;
; Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 60
DB 6 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 65
QY 61 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 100

RESULT 9
US-09-056-556-214
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; QUERY MATCH
; BEST LOCAL SIMILARITY 100.0%; Score 466; DB 4; Length 100;
; Mismatches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Query Match 100.0%; Score 466; DB 4; Length 100;
; Best Local Similarity 100.0%; Pred. No. 2.3e-46;
; Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 60
DB 6 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEANKQ 65
QY 61 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 95
DB 66 KOELDEISTNIRQAGVOYSRADDEQQQALSSQMGF 100

TREATM
```


Db 489 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 523

RESULT 12
US-09-116-492A-5
; Sequence 5, Application US/09116492A
; Patent No. 6436409
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOB
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492A-5

Query Match 99.1%; Score 462; DB 4; Length 100;
Best Local Similarity 98.9%; Pred. No. 6.5e-46;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDAATLAQEAQNERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60
DB 6 TDAATLQEAQNERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 65

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
DB 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 13
US-08-818-112-117
; Sequence 117, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 622-4900
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-117

Query Match 84.1%; Score 392; DB 4; Length 80;

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-117

Query Match 84.1%; Score 392; DB 3; Length 80;
Best Local Similarity 100.0%; Pred. No. 5.3e-38;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQKQELDEISTNIRQAG 75
DB 1 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQKQELDEISTNIRQAG 60

QY 76 VOYSRADEEQQALSSQMGF 95
DB 61 VOYSRADEEQQALSSQMGF 80

RESULT 14
US-08-818-111-112
; Sequence 112, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-112

Query Match 84.1%; Score 392; DB 4; Length 80;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 3.70522 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 PDATLAQENGRISGL.....VOYSADEEQOALLSSQWGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US40_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US80_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	95	12	US-10-084-843-88
2	466	100.0	95	12	US-10-193-002-89
3	466	100.0	95	12	US-10-098-732A-35
4	466	100.0	100	12	US-10-084-843-115
5	466	100.0	100	12	US-10-193-002-110
6	466	100.0	100	16	US-10-080-170-639
7	466	100.0	358	9	US-09-287-849-8
8	466	100.0	358	12	US-10-359-460-8
9	466	100.0	802	9	US-09-287-849-10
10	466	100.0	802	12	US-10-084-843-214
11	466	100.0	802	12	US-10-084-843-351
12	466	100.0	802	12	US-10-193-002-209
13	466	100.0	802	12	US-10-193-002-346
14	466	100.0	802	12	US-10-359-460-10
15	462	99.1	100	15	US-10-140-045-5

16	392	84.1	80	12	US-10-084-843-117	Sequence 117, App
17	392	84.1	80	12	US-10-193-002-112	Sequence 112, App
18	217	46.6	49	15	US-10-140-045-6	Sequence 6, Appl
19	205	44.0	42	15	US-10-140-045-7	Sequence 7, Appl
20	173	37.1	100	15	US-10-140-045-28	Sequence 28, Appl
21	173	37.1	100	16	US-10-080-170-12	Sequence 12, Appl
22	137	29.4	28	15	US-10-140-045-12	Sequence 12, Appl
23	122	26.2	28	12	US-10-084-843-98	Sequence 98, Appl
24	121	26.0	27	12	US-10-084-843-96	Sequence 96, Appl
25	118	25.3	27	12	US-10-084-843-95	Sequence 95, Appl
26	118	25.3	27	12	US-10-084-843-97	Sequence 97, Appl
27	117	25.1	28	12	US-10-084-843-93	Sequence 93, Appl
28	100.5	21.6	1202	15	US-10-156-761-10195	Sequence 10195, A
29	100	21.5	20	15	US-10-140-045-8	Sequence 8, Appl
30	91	19.5	108	10	US-09-738-626-4144	Sequence 4144, Ap
31	87.5	18.8	97	9	US-09-073-009-109	Sequence 109, App
32	87.5	18.8	97	9	US-09-073-009-143	Sequence 143, App
33	87.5	18.8	97	9	US-09-793-306-109	Sequence 109, App
34	87.5	18.8	97	9	US-09-793-306-143	Sequence 143, App
35	87.5	18.8	97	12	US-10-098-732A-24	Sequence 24, Appl
36	87.5	18.8	299	9	US-09-287-849-19	Sequence 19, Appl
37	87.5	18.8	299	12	US-10-359-460-19	Sequence 19, Appl
38	87.5	18.8	299	12	US-10-098-732A-47	Sequence 47, Appl
39	87.5	18.8	433	9	US-09-287-849-14	Sequence 14, Appl
40	87.5	18.8	433	12	US-10-359-460-14	Sequence 14, Appl
41	87.5	18.8	710	9	US-09-287-849-16	Sequence 16, Appl
42	87.5	18.8	710	12	US-10-359-460-16	Sequence 16, Appl
43	87.5	18.8	710	12	US-10-098-732A-49	Sequence 49, Appl
44	87.5	18.8	856	9	US-09-287-849-12	Sequence 12, Appl
45	87.5	18.8	856	12	US-10-359-460-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-10-084-843-88
; Sequence 88, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

```
;
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-10-084-843-88
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 2
US-10-193-002-89
; Sequence 89, Application US/10193002
; Publication No. US2003013502A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; SKEIKY, YASIR A.W.
; DILLON, DAVIN C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-10-193-002-89
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 3
US-10-098-732A-35
; Sequence 35, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Tb38-1 or 38-1 (MTb11)
US-10-098-732A-35
Query Match 100.0%; Score 466; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.9e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60
Db 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGAGTAAGAAVVRFOEAANKQ 60

Qy 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95
Db 61 KOELDEISTNIRQAGVOYSRADEEQQALSSQMGF 95

RESULT 4
US-10-084-843-115
; Sequence 115, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; SKEIKY, YASIR A.W.
; DILLON, DAVIN C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
```


AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 115:

US-10-084-843-115

Query Match 100.0%; Score 466; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy. 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQQALLSSQMGF 95
Db 66 KOELDEISTNIRQAGVQYSRADDEQQQALLSSQMGF 100

RESULT 5
US-10-193-002-110
Sequence 110, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 110:

US-10-193-002-110

Query Match 100.0%; Score 466; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQQALLSSQMGF 95
Db 66 KOELDEISTNIRQAGVQYSRADDEQQQALLSSQMGF 100

RESULT 6
US-10-080-170-639
Sequence 639, Application US/10080170
Publication No. US20030129601A1
GENERAL INFORMATION:
APPLICANT: COLE, S.T.
TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIAL GENOMICS AS A TOOL FOR
TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170
CURRENT FILING DATE: 2002-06-10
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 652
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 639
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis

US-10-080-170-639

Query Match 100.0%; Score 466; DB 16; Length 100;
Best Local Similarity 100.0%; Pred. No. 2e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFRISGLTKTQIDQVESTAGSLQGWGGAAGTAAGAAVVRFOEAANKQ 65

Qy 61 KOELDEISTNIRQAGVQYSRADDEQQQALLSSQMGF 95

```
Db 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
|||||
PRT
RESULT 7
US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8
Query Match 100.0%; Score 466; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 9.6e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 60
|||||
Db 264 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 323
|||||
QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
|||||
Db 324 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 358
|||||
RESULT 8
US-09-287-849-8
; Sequence 8, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 2003-02-05
```

```
; PRIOR APPLICATION NUMBER: US 09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TbH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-359-460-8
Query Match 100.0%; Score 466; DB 12; Length 358;
Best Local Similarity 100.0%; Pred. No. 9.6e-43;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 60
|||||
Db 264 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 323
|||||
QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
|||||
Db 324 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 358
|||||
RESULT 9
US-09-287-849-10
; Sequence 10, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10

Query Match 100.0%; Score 466; DB 9; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 60
DB 429 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 488

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 523

RESULT 10
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 214:
US-10-084-843-214

Query Match 100.0%; Score 466; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 60

DB 429 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 488

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 523

RESULT 11
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 351:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match 100.0%; Score 466; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 60
DB 429 TDAATLAQAGNFERISGDLTKTIDQVESTAGSLQGWGGAAGTAAGAAQAAVVRFOEAANKQ 488

QY 61 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 95
DB 489 KOELDEISTNIRQAGVQYSRADDEEQOQALSSQMGF 523

RESULT 12

US-10-193-002-209
 ; Sequence 209, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedvick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS

NUMBER OF SEQUENCES: 350
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/193.002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/072.596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 209:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 802 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-10-193-002-209
 Query Match 100.0%; Score 466; DB 12; Length 802;
 Best Local Similarity 100.0%; Pred. No. 2.6e-42;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 DB 429 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 488

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 DB 489 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 523

RESULT 13

US-10-193-002-346
 ; Sequence 346, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.

Dillon, Davin C.
 Campos-Neto, Antonia
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS

NUMBER OF SEQUENCES: 350
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/193.002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/072.596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 346:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 802 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 346:
 US-10-193-002-346

Query Match 100.0%; Score 466; DB 12; Length 802;
 Best Local Similarity 100.0%; Pred. No. 2.6e-42;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
 DB 429 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 488

QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
 DB 489 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 523

RESULT 14

US-10-359-460-10
 ; Sequence 10, Application US/10359460
 ; Publication No. US20030147911A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Corixa Corporation

TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

TITLE OF INVENTION: and Their Uses

FILE REFERENCE: 014059-009020US

Db 66 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100
Search completed: November 21, 2003, 16:38:12
Job time : 4.70522 secs

Query Match 100.0%; Score 466; DB 12; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.6e-42;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60
Db 429 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 488
QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95
Db 489 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 523

RESULT 15
US-10-140-045-5
Sequence 5, Application US/10140045
Publication No. US20030092899A1
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
APPLICANT: BERTHET, FRANCOIS-XAVIER
APPLICANT: ANDERSEN, PETER
APPLICANT: RASMUSSEN, PETER B
TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS VECTORS AND METHODS FOR THE PRODUCTION OF THE SAME
TITLE OF INVENTION: USING THE SAME
FILE REFERENCE: 0660-0137-27X
CURRENT APPLICATION NUMBER: US/10/140,045
CURRENT FILING DATE: 2002-05-08
PRIOR APPLICATION NUMBER: US/09/116,492A
PRIOR FILING DATE: 1998-07-16
PRIOR APPLICATION NUMBER: 60/252,631
PRIOR FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 100
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-10-140-045-5

Query Match 99.1%; Score 462; DB 15; Length 100;
Best Local Similarity 98.9%; Pred. No. 5.5e-43;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 60
Db 6 TDAATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAANKQ 65
QY 61 KQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 95

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 1.98186 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQAGNFRISGDL.....VOYSGRADEQQQALLSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	100	2 H70802	hypothetical prote
2	173	37.1	100	2 T10032	hypothetical prote
3	93.5	20.1	1186	2 T35661	probable chromosom
4	90.5	19.4	410	2 B59103	hypothetical prote
5	90	19.3	750	2 T38435	coiled coil protei
6	87.5	18.8	97	2 B70836	probable transcrip
7	86	18.5	103	2 B70600	hypothetical prote
8	86	18.5	2022	2 T43244	ovtl protein - nem
9	84.5	18.1	97	2 G70857	probable PE protei
10	81.5	17.5	955	2 A47334	Lckin kinesin-rela
11	80	17.2	527	2 S33068	myosin heavy chain
12	80	17.2	1940	2 A59287	myosin heavy chain
13	79.5	17.1	245	2 T44704	hypothetical prote
14	79.5	17.1	247	2 G87116	conserved hypochet
15	79.5	17.1	1361	1 A61231	myosin heavy chain
16	79.5	17.1	1999	1 S1801	myosin heavy chain
17	79	17.0	508	1 K5SH12	keratin type II, m
18	78	16.7	1325	2 T42722	male-enhanced anti
19	76	16.3	491	2 S05408	keratin type II,
20	76	16.3	2094	2 S33124	tpi protein - huma
21	75	16.1	163	2 P75559	hypothetical prote
22	74.5	16.0	80	2 D29674	phycocyanin linker
23	74.5	16.0	80	2 A61873	rod-capping linker
24	74.5	16.0	316	2 JC6549	apolipoprotein E p
25	74.5	16.0	547	2 H82503	conserved hypochet
26	74	15.9	528	2 B75310	methyl-accepting c
27	74	15.9	2101	2 A42184	nuclear mitotic ap
28	73.5	15.8	433	2 D84335	hypothetical prote
29	73.5	15.8	440	1 S33479	ferredoxin-NADP re

30 73.5 15.8 440 2 AB0321 ferredoxin-NADP(+) 31 73.5 15.8 552 2 AD1824 periplasmic oligop 32 73.5 15.8 745 2 H96760 unknown protein [i 33 73.5 15.8 1313 2 F96673 hypothetical prote 34 73.5 15.8 1679 2 S48385 hypothetical prote 35 73 15.7 103 2 G96905 uncharacterized sm 36 73 15.7 257 2 I38025 keratin-like prote 37 73 15.7 1238 2 T03465 probable exonuclea 38 72.5 15.6 80 2 A24691 8.9k linker polype 39 72 15.5 503 2 S29094 keratin, type II, 40 71.5 15.3 293 1 DAPSPC biphényl-2,3-diol 41 71.5 15.3 392 2 G95258 secreted 45 kd pro 42 71.5 15.3 392 2 B98124 general stress pro 43 71.5 15.3 555 2 D87536 methyl-accepting c 44 71.5 15.3 1170 2 A53612 laminin B1k chain 45 71.5 15.3 1492 2 T14652 protein J - Yersin

ALIGNMENTS

RESULT 1

H70802

hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: H70802

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R/Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, R.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: H70802

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-100 <COL>

A/Cross-references: GB:AL022120; GB:AL123456; NID:G3261558; PIDN:CAA17966.1; PID:G296022

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV3874

Query Match 100.0%; Score 466; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 9e-36;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 60
|||||
Db 6 TDAATLAQAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAQAQAAVVFQEAANKQ 65
Qy 61 KQELDISTRIRQAGVOYSGRADEEQQALLSSQMGF 95
|||||
Db 66 KQELDISTRIRQAGVOYSGRADEEQQALLSSQMGF 100

RESULT 2

T10032

hypothetical protein MLCB28.13c - Mycobacterium leprae

C/Species: Mycobacterium leprae

C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000

C/Accession: T10032

R/Higmeier, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.

Mol. Microbiol. 7, 197-206, 1993

A/Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobact

A/Reference number: Z16917; MUID:93188700; PMID:8446027

A/Accession: T10032

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-100 <BIG>

A/Cross-references: EMBL:Y14967; NID:G2370268; PIDN:CAA75210.1; PID:G2370280

Query Match 37.1%; Score 173; DB 2; Length 100;

hypothetical protein RV3805c - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: B70600
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: B70600
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-103 <COL>
A/Cross-references: GB:Z94121; GB:AL123456; NID:G3261736; PIDN:CAB08096.1; PID:e312272;
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: RV3805c

Query Match 18.5%; Score 86; DB 2; Length 103;
Best Local Similarity 23.9%; Pred. No. 0.38;
Matches 21; Conservative 20; Mismatches 47; Indels 0; Gaps 0;
QY 2 DAATLAQAGNFERISGLKTCIQDQVESTAGSLQGWEGAGTAQAQAVRFOEAAANKQX 61
Db 10 EPVAVQGGFAALDGAHEHLAVQLAELDAQVGMGGWKGASGAYSAGWELWHRGAGEVQ 69
QY 62 QELDEISTNIRQAGVQYGRABEQOQAL 89
Db 70 LGLSLMAAAIAHAGAGYQHNETASQVL 97

RESULT 8
T43214
Ov11 protein - nematode (Onchocerca volvulus)
A/Alternate names: myosin-like antigen
C/Species: Onchocerca volvulus
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 20-Jun-2000
C/Accession: T43214; A44939; A54513; S27825
R/Rittertrapapab, S.; Richie, T.L.; Tuan, R.S.; Shepley, K.J.; Dinman, J.D.; Neubert, T.
Mol. Biochem. Parasitol. 69, 161-171, 1995
A/Title: Molecular cloning of a gene expressed during early embryonic development in Onchocerca volvulus
A/Reference number: Z22341; MUID:95287898; PMID:7770081
A/Accession: T43214
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2022 <TRI>
A/Cross-references: EMBL:U12681; NID:G530824; PID:G530825; PIDN:AAA80009.1
A/Experimental source: specific host Homo sapiens
R/Erond, N.E.; Donelson, J.E.
Mol. Biochem. Parasitol. 40, 213-224, 1990
A/Title: Characterization of a myosin-like antigen from Onchocerca volvulus.
A/Reference number: A44939; MUID:90301142; PMID:2194123
A/Accession: A44939
A/Molecule type: mRNA
A/Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1363 <ERO>
A/Cross-references: GB:M30398
A/Note: the sequence is revised in GenBank entry ONCANTML, release 115, (PIDN:AAA29413.1
R/Donelson, J.E.; Duke, S.O.B.; Moser, D.; Zeng, W.; Erond, N.E.; Lucius, R.; Renz, A.
Mol. Biochem. Parasitol. 31, 241-250, 1988
A/Title: Construction of Onchocerca volvulus cDNA libraries and partial characterization
A/Reference number: A54513; MUID:89127417; PMID:2464764
A/Accession: A54513
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 733-866 <DON>
A/Cross-references: GB:J03095; NID:G159874; PIDN:AAA29412.1; PID:G159875
R/Ngazi, E.; Erond, N.E.; Donelson, J.E.
submitted to the EMBL Data Library, April 1990
A/Description: Characterization of a myosin-like antigen from Onchocerca volvulus.
A/Reference number: S27825
A/Accession: S27825

A/Molecule type: mRNA
A/Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1347, 'E', 1
A/Cross-references: EMBL:M30398; NID:G159876; PIDN:AAA29413.1; PID:G159877
C/Genetics:
A/Gene: Ov11
C/Keywords: leucine zipper
Query Match 18.5%; Score 86; DB 2; Length 2022;
Best Local Similarity 25.5%; Pred. No. 9.8;
Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps 2;
QY 6 LAQAGNFERISGLKTCIQDQVESTAGSLQGWEGAGTAQAQAVRFOEAAANKQX 55
Db 1752 LDEKRTMENILHETALQREATIESSLNALERENKELHRNCAQLQOQIAQLENGNRLIQ 1811
QY 56 AANKQOQELDEISTNIRQAGVQYGRABEQOQALSSOM 93
Db 1812 LTNKQREYDKFAQNWRTKIQIERIENRSLKRSRI 1849

RESULT 9
G70857
Probable PE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: G70857
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: G70857
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-97 <COL>
A/Cross-references: GB:AL021287; GB:AL123456; NID:G3261508; PIDN:CAA16105.1; PID:e123776
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: PE
Query Match 18.1%; Score 84.5; DB 2; Length 97;
Best Local Similarity 34.9%; Pred. No. 0.49;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;
QY 13 FERISGLKTCIQDQVESTAGSLQGWEGAGTAQAQAVRFOEAAANKQOQELDEISTNIR 72
Db 18 FFAKAGLMRHTICQAEQQAMSAQAQPHQGSAAAFQGAHARFVAAAANKVNTLIDIAQANLG 77
QY 73 QAGVQYGRABEQOQALSSOMGF 95
Db 78 EAAGTTVAAD---AAAASYTGF 97

RESULT 10
A47334
LcKin kinesin-related antigen - Leishmania chagasi (fragment)
C/Species: Leishmania chagasi
C/Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001
C/Accession: A47334
R/Burns Jr., J.M.; Shreffler, W.G.; Benson, D.R.; Ghalib, H.W.; Badaro, R.; Reed, S.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 775-779, 1993
A/Title: Molecular characterization of a kinesin-related antigen of Leishmania chagasi
A/Reference number: A47334; MUID:93133867; PMID:8421715
A/Accession: A47334
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-955 <BUR>
A/Cross-references: GB:I07879; NID:G308884; PIDN:AAA29254.1; PID:G308885
A/Experimental source: MHOM/BS/82/BA-2, Cl
A/Note: sequence extracted from NCBI backbone (NCBI:122864, NCBI:122865)
C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; nucleotide binding; P-loop
F;13-398/Domain: kinesin motor domain homology <KMOT>
F;122-129/Region: nucleotide-binding motif A (P-loop)

Query Match 17.5%; Score 81.5; DB 2; Length 955;
Best Local Similarity 31.0%; Pred. No. 11;
Matches 31; Conservative 12; Mismatches 44; Indels 13; Gaps 2;

QY 4 ATLAQAGNFRISGDLKTDQV-----ESTAGSLQGWGAGTAAGTAQAQAVR 52
DB 858 ATLEQLRSEERAEALASOLETTAAKMSAEQDRESTRATLEQLRDSERAEALASOL 917

QY 53 FOEAANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQ 92

DB 918 EATAAKSAEQDR--ENTRAALEQLRDSERAEALASQ 955

RESULT 11

S33068 myosin heavy chain - fluke (Schistosoma mansoni) (fragment)

N;Alternate names: surface antigen, 200K

C;Species: Schistosoma mansoni

C;Date: 22-Nov-1993 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998

C;Accession: S33068

R;Seiison, L.M.A.; Masterson, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.

J. Immunol. 149, 3612-3620, 1992

A;Title: Induction of protective immunity in mice using a 62-kDa recombinant fragment of

A;Reference number: A46514; MUID:93056536; PMID:1431131

A;Accession: S33068

A;Molecule type: mRNA

A;Residues: 1-527 <SOI>

A;Cross-references: EMBL:X65591

A;Note: the authors translated the codon CAA for residue 346 as Lys

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: ATP; surface antigen

Query Match 17.2%; Score 80; DB 2; Length 527;

Best Local Similarity 22.8%; Pred. No. 8;

Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;

QY 2 DAATLAQAGNFRISGDLKTDQVSTAGSLQGWGAGTAAGTAQAQAVRFOEAANKQK 61

DB 341 DLKATQETVDDLRRVRLDEQLRRKEAIGGLSGKFEDEQGLVAQ-----LQRIKELQ 395

QY 62 QELDEISTNIRQAGVQYSRADDEQQQALSSQ 93

DB 396 TRIQEELEDEAPRAARSKAKSRQKQ-LESEL 426

RESULT 12

A59287

myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)

C;Species: Schistosoma mansoni

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000

C;Accession: A59287

R;Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.

Mol. Biochem. Parasitol. 58, 161-164, 1993

A;Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA from

A;Reference number: A59287; MUID:93211444; PMID:8459827

A;Accession: A59287

A;Status: Preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1940 <WES>

A;Cross-references: GB:L01634; PIDN:AAA29905.1

A;Experimental source: strain Brazilian LE

C;Genetics:

A;Gene: MYH

C;Superfamily: myosin heavy chain; myosin motor domain homology

F;82-752/Domain: myosin motor domain homology <MMO>

Query Match 17.2%; Score 80; DB 2; Length 1940;

Best Local Similarity 22.8%; Pred. No. 33;

Matches 21; Conservative 21; Mismatches 44; Indels 6; Gaps 2;

QY 2 DAATLAQAGNFRISGDLKTDQVSTAGSLQGWGAGTAAGTAQAQAVRFOEAANKQK 61
DB 1044 DLKATQETVDDLRRVRLDEQLRRKEAIGGLSGKFEDEQGLVAQ-----LQRIKELQ 1098

QY 62 QELDEISTNIRQAGVQYSRADDEQQQALSSQ 93

DB 1099 TRIQEELEDEAPRAARSKAKSRQKQ-LESEL 1129

RESULT 13

T44704

hypothetical protein MLCB1243.13 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000

C;Accession: T44704

R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1998

A;Reference number: 222830

A;Accession: T44704

A;Status: Preliminary; translated from GB/EMBL/DDBU

A;Molecule type: DNA

A;Residues: 1-245 <PAR>

A;Cross-references: EMBL:AL023635; PIDN:CAA19194.1

A;Experimental source: cosmid B1243

C;Genetics:

A;Note: MLCB1243.13

C;Superfamily: Mycobacterium leprae hypothetical protein MLCB1243.13

Query Match 17.1%; Score 79.5; DB 2; Length 245;

Best Local Similarity 27.0%; Pred. No. 3.9;

Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNFERISGDLKTDQVSTAGS-----LOGQWRGAAGTAQAQAVRFOE 55

DB 79 TESESLSHARAEADRIILSDAKSQVDRMASEAQHSERMLGDRARESIATVAKREYEA 138

QY 56 AANKQKQELDEISTNIRQAGVQYSRADDE---QQQALSSQ 92

DB 139 SLNRAQSECDRL---IENGNI SYEKAIQEGIKEQQLVLSQ 175

RESULT 14

GS7116

conserved hypothetical protein MLI661 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C;Accession: GS7116

R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A;Title: Massive gene decay in the leprosy bacillus

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: GS7116

A;Status: Preliminary

A;Molecule type: DNA

A;Residues: 1-247 <STO>

A;Cross-references: GB:AL450380; MID:g13093432; PIDN:CAC30614.1; GSPDB:GN00147

C;Genetics:

A;Gene: MLI661

C;Superfamily: Mycobacterium leprae hypothetical protein MLCB1243.13

Query Match 17.1%; Score 79.5; DB 2; Length 247;

Best Local Similarity 27.0%; Pred. No. 3.9;

Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNFERISGDLKTDQVSTAGS-----LOGQWRGAAGTAQAQAVRFOE 55

DB 81 TESESLSHARAEADRIILSDAKSQVDRMASEAQHSERMLGDRARESIATVAKREYEA 140

QY 56 AANKQKQELDEISTNIRQAGVQYSRADDE---QQQALSSQ 92

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 1.12018 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQEAQNFERISGDL.....VOYGRADEQQQALSSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	466	100.0	99	CF10_MYCTU	O69739 mycobacteri
2	173	37.1	99	CF10_MYCTU	O33084 mycobacteri
3	86	18.5	103	ES6D_MYCTU	O05440 mycobacteri
4	86	18.5	2022	ANT1_ONCVO	P21249 onchocerca
5	81.5	17.5	955	KINL_LEICH	P46865 leishmania
6	79.5	17.1	245	Y127_MYCLE	O9cbs6 mycobacteri
7	79.5	17.1	1960	MYH9_HUMAN	P35579 homo sapien
8	78	16.7	1325	G160_MOUSE	P35937 mus musculu
9	76	16.3	491	K2M2_SHEEP	P15241 ovis aries
10	76	16.3	2349	TPR_HUMAN	P12270 homo sapien
11	75.5	16.2	440	FENR_ANAVA	Q44549 anabaena va
12	75.5	16.2	1961	MYH9_RAT	Q62812 rattus norv
13	74.5	16.0	80	PYS1_ANASP	P07124 anabaena sp
14	73.5	15.8	440	FENR_ANASP	P21890 anabaena sp
15	73.5	15.8	440	FENR_ANASP	P35558 anabaena sp
16	73.5	15.8	1679	Y109_YEAST	P40457 saccharomyc
17	73	15.7	1238	SBCR_RHOCA	O68032 rhodobacter
18	72.5	15.6	80	PYS1_MASLA	P11396 mastigoclad
19	72.5	15.6	1947	MYSC_CABEL	P21845 caenorhabdi
20	72	15.5	502	K2M3_SHEEP	P25691 ovis aries
21	71.5	15.3	292	BPC3_PSES1	P17297 pseudomonas
22	71.5	15.3	1172	LMB3_HUMAN	Q13751 homo sapien
23	71	15.2	1938	MYSD_CABEL	P02567 caenorhabdi
24	70.5	15.1	78	PYS1_SYNEL	P05035 synechococc
25	70.5	15.1	189	APL3_MANGE	P13276 manduca sex
26	70.5	15.1	373	CK13_HUMAN	Q02833 homo sapien
27	70.5	15.1	845	Y4PA_RHISN	P55439 rhizobium s
28	70.5	15.1	1959	MYH9_CHICK	P54105 gallus gall
29	70	15.0	186	APL3_GALME	P80703 gallus me
30	70	15.0	512	MGPD_ENTAE	P21823 enterobacteri
31	69.5	14.9	242	GRPE_HALME	Q9hc2 halobacteri
32	69.5	14.9	457	MESE_LEUME	Q10419 leuconostoc
33	69	14.8	232	YM29_MYCLE	O9cbs9 mycobacteri

RESULT 1

CF10_MYCTU	STANDARD;	FRT;	99 AA.
AC	O69739;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	10 kDa culture filtrate antigen cfp10 (Secreted antigenic protein MTSA-10).		
DE	MTSA-10.		
GN	CFP10 OR LHP OR MTSA10 OR RV3874 OR MT3988 OR MTV027.09.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
CX	NCBI_TaxID=1773;		
[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.		
RC	STRAIN=H37RV;		
RX	MEDLINE=99061212; PubMed=9846755;		
RA	Berthet P.-X., Rasmussen P.B., Rosenkrands I., Andersen P., Gicquel B.;		
RT	"A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10).";		
RL	Microbiology 144:3195-3203(1998).		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia P., Davies R., Devlin K., Felwell T., Chillingworth T., Connor R., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Oliver S., Osborne J., Quail M.A., McLean J., Moule S., Murphy L., Hornsby T., Jagels K., Krogh A., Rajandream M.A., Rogers J., Sulston J.E., Taylor K., Whitehead S., Squares R., Rutter S., Seeger K., Skelton S., Squares S., Squares R., "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";		
RL	Nature 393:537-544(1998).		
[3]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Emdinova M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
[4]			
RP	SEQUENCE FROM N.A.		
RA	Singh B., Siddiqui Z., Singh S., Sharma P.;		
RT	"rv3874 (mtsa-10) gene of a clinical isolate of Mycobacterium tuberculosis from India.";		
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: Secreted.		

ALIGNMENTS

P48667 homo sapien
P33741 halobacteri
P48668 homo sapien
Q10973 mycobacteri
P12008 synchococc
P21258 bos taurus
P33292 pichia past
Q9nqx4 homo sapien
Q8bx73 homo sapien
Q8xg13 salmonella
P13509 alcaligenes
P06159 human parai

34 69 14.8 384 1 K2CD_HUMAN
35 69 14.8 535 1 HTR1_HALN1
36 69 14.8 563 1 K2CE_HUMAN
37 68.5 14.7 245 1 Y127_MYCTU
38 68.5 14.7 356 1 IRPA_SYNP7
39 68.3 14.7 453 1 MSRE_BOVIN
40 68.5 14.7 576 1 PEX5_PICPA
41 68.5 14.7 1742 1 MY5C_HUMAN
42 68.5 14.7 2442 1 CEP2_HUMAN
43 68 14.6 327 1 SYFA_SALTY
44 68 14.6 417 1 CZCC_ALCEU
45 68 14.6 515 1 NCAP_P13H4

CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004671; AAC83445.1; -
CC EMBL; AL022120; CAA17966.1; -
CC EMBL; AE007190; AAK48356.1; -
CC EMBL; AF419854; AAL14999.1; -
CC DR PIR; H70802; H70802.
CC DR TIGR; MT3988; -
CC DR Tuberculin; Rv3874; -
CC KW Antigen; Complete proteome.
CC FT INIT MET 0
CC SQ SEQUENCE 99 AA; 10663 MW; EBCAE6A996C5489D CRC64;

Query Match 100.0%; Score 466; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 5.3e-36;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 5 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 64

QY 61 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 95
Db 65 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 99

RESULT 2
CF10_MYCLE STANDARD; PRT; 99 AA.
AC O33084;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 10 kDa culture filtrate antigen cf10 homolog.
GN MLC050 OR MLCB628.13C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.,
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RN complete genome sequence."
RN Nature 393:537-544 (1998).
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RN Nature 409:1007-1011 (2001).
RP SEQUENCE FROM N.A.
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y14967; CAA75210.1; -
CC EMBL; AL583917; CAC29958.1; -
CC DR PIR; T10032; T10032.
CC DR Leproma; MLC050; -
CC KW Complete proteome.
CC FT INIT MET 0
CC SQ SEQUENCE 99 AA; 10833 MW; 859B484F7E58A8A CRC64;

Query Match 37.1%; Score 173; DB 1; Length 99;
Best Local Similarity 37.9%; Pred. No. 2.3e-09;
Matches 36; Conservative 20; Mismatches 39; Indels 0; Gaps 0;

QY 1 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60
Db 5 TDAATLAQEAAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 64

QY 61 KOELDEISTNIRQAGVOYSRADEEQQAALSSQMGF 95
Db 65 IRQLESIVDKLNRSGNYTKTDDANQLLSKMF 99

RESULT 3
ES6D_MYCTU STANDARD; PRT; 103 AA.
ID ES6D_MYCTU
AC O05440;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ESAT-6 like protein 13.
GN RV3905C OR Mt4024 OR MTCY15F10.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RN complete genome sequence."
RN Nature 393:537-544 (1998).
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
DR EMBL; Z94121; CAB08096.1; .....
DR EMBL; ABC07193; AAK48388.1; .....
DR PIR; B70600; B70600.
DR TIGR; MT4024; .....
KW Tuberculin; Bv3905c; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 10460 MW; 3994E272A7BDF02 CRC64;

Query Match      18.5%; Score 86; DB 1; Length 103;
Best Local Similarity 23.9%; Pred. No. 0.2;
Matches 21; Conservative 20; Mismatches 47; Indels 0; Gaps 0;

QY 2 DAATLAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGCTAAQAQAAVVFQEAANKOK 61
Db 10 EFVAVQGFPAASLDGAHHLAVGLAELDAQVGMQLGCGWAGSAYGSAWELWHRGAGEVQ 69
QY 62 QELDEISTNIRQAGVQYSRADEEQOQAL 89
Db 70 LGLSMLAAIAHAGAGYQHNETASQVL 97

RESULT 4
ANTI_ONCVO
ID ANTI_ONCVO STANDARD; PRT; 2022 AA.
AC P21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-NOV-2000 (Rel. 39, Last annotation update)
DE Major antigen.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95287898; PubMed=7770081;
RA Triteaxaprab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA Neubert T.A., Scott A.L.;
RT "Molecular cloning of a gene expressed during early embryonic
development in Onchocerca volvulus."
RL Mol. Biochem. Parasitol. 69:161-171(1995).
RN [2]
SEQUENCE OF 733-866 FROM N.A.
RX MEDLINE=89127417; PubMed=2464764;
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erondou N.E.,
RA Lucius R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
characterization of the cDNA for a major antigen."
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U12681; AAA80009.1; .....
DR EMBL; J03995; AAA29412.1; .....
DR PIR; T43214; T43214.
KW Antigen; Coiled coil.
FT DOMAIN 74 120 COILED COIL (POTENTIAL).
FT DOMAIN 151 251 COILED COIL (POTENTIAL).
FT DOMAIN 327 384 COILED COIL (POTENTIAL).
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match      18.5%; Score 86; DB 1; Length 2022;
Best Local Similarity 25.5%; Pred. No. 4.7;
Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps 2;

QY 6 LAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGCTAAQAQAAVVFQEAANKOK 55
Db 1752 LDEKRTMENILHETALQRAIESSNALARENKELHRNCAQLQQQIAQLENGNELIQ 1811
QY 56 AANKQOELDEISTNIRQAGVQYSRADEEQOQALSSQM 93
Db 1812 LTNKQREYDKFAQNVRTXIQIETIENRSLKSGRI 1849

RESULT 5
KINL_LEICH
ID KINL_LEICH STANDARD; PRT; 955 AA.
AC P46865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein K39 (Fragment).
GN KIN.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=MEOM/BR/82 / Isolate BA-2;
RX MEDLINE=93133867; PubMed=8421715;
RA Burns J.M. Jr., Shreffler W.G., Benson D.R., Ghalib H.W., Badaro R.,
RA Reed S.G.;
RT "Molecular characterization of a kinesin-related antigen of
Leishmania chagasi that detects specific antibody in African and
American visceral leishmaniasis."
RL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC -!- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L07879; AAA29254.1; .....
DR PIR; A47334; A47334.
DR HSP; P17119; 3KAR.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SMO0129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Repeat.
FT DOMAIN 1 399 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 426 >955 COILED COIL (POTENTIAL).
FT NP_BIND 122 129 ATP (POTENTIAL).
FT DOMAIN 704 >955 7 X 39 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 704 742 1.
FT REPEAT 743 781 2.
FT REPEAT 782 820 3.
FT REPEAT 821 859 4.
FT REPEAT 860 898 5.
FT REPEAT 899 937 6.
FT REPEAT 938 >955 7 (PARTIAL).
FT NON_TER 955 955
SQ SEQUENCE 955 AA; 106168 MW; 8CA76815BE84C6E9 CRC64;

Query Match      17.5%; Score 81.5; DB 1; Length 955;
Best Local Similarity 31.0%; Pred. No. 5.4;
Matches 31; Conservative 12; Mismatches 44; Indels 13; Gaps 2;

QY 4 ATLAQEAAGNERISGDLTKTQIDQVESTAGSLQGWGAGCTAAQAQAAVVFQEAANKOK 52
```

```

D5 858 ATLEQQRESEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRDSEERAAELASQ 917
QY 53 FQEAANKQKQELDEISNIRQAGVQVSRADDEEQQALSSQ 92
D5 918 EATRAAKSAEQDR--ENTRAALEQQLRDSEERAAELASQ 955

RESULT 6
YT27 MYCLE
ID YT27 MYCLE STANDARD; PRT; 245 AA.
AC Q9CB86; 069467;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein MLI1661.
GN MLI1661 OR MLCB1243.13.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2927C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL023635; CAAL19194.1;
DR EMBL; AL383923; CAC30614.1; ALT_INIT.
DR PIR; T44704; T44704.
DR Leproma; MLI1661; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 245 AA; 27087 MW; C984D9A5FA49697A CRC64;

Query Match 17.18; Score 79.5; DB 1; Length 245;
Best Local Similarity 27.08; Pred. No. 1.9;
Matches 27; Conservative 20; Mismatches 42; Indels 11; Gaps 4;

QY 1 TDAATLAQEA-GNPERISGLTKTQIDQVESTAGS----LOGQWRGAAGTAAQAQAVRFOE 55
D5 79 TESESLSHARAEADRLSDAKSQVDRMASEARQHSERMLGDAREBSIRIATVAKREYEA 138
QY 56 AANKQKQELDEISNIRQAGVQVSRADDEEQQALSSQ 92
D5 139 SLNPAQSECRL-----IENGNISVEKATQEGIKQQLVLSQ 175

RESULT 7
MYH9_HUMAN
ID MYH9_HUMAN STANDARD; PRT; 1960 AA.
AC P35579; 060805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 28-FEB-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,

```

```

DE GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beate D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.P., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Evans P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Dhami K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Gratham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashregui-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsay Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Swann R.M.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.B., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dunanski J.P., Beyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [2]
RP SEQUENCE OF 1-1337 FROM N.A.
RC MEDLINE=92000925; PubMed=1912569;
RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
RA Arnaout M.A., Clayton L.K., Tenen D.G.;
RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
RT cDNA clones, characterization of the protein, chromosomal
RT localization, and upregulation during myeloid differentiation.";
RL Blood 78:1826-1833(1991).
RN [3]
RP SEQUENCE OF 1-715 FROM N.A.
RC MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes";
RL Circ. Res. 69:530-539(1991).
RN [4]
RP SEQUENCE OF 714-1960 FROM N.A.
RC MEDLINE=90138958; PubMed=1967836;
RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;

```


DR	MIM; 153640;	-
DR	MIM; 155100;	-
DR	MIM; 603622;	-
DR	MIM; 605249;	-
DR	GO; GO:0005860; C:non-muscle myosin; TAS.	
DR	InterPro; IPR000048; IQ region.	
DR	InterPro; IPR001609; myosin head.	
DR	InterPro; IPR004009; Myosin N.	
DR	InterPro; IPR002928; Myosin tail.	
DR	InterPro; IPR002017; Spectrin.	
DR	Pfam; PF00612; IQ; 1.	
DR	Pfam; PF00663; myosin head; 1.	
DR	Pfam; PF02736; Myosin_N; 1.	
DR	Pfam; PF01576; Myosin_tail; 1.	
DR	PRINTS; PR00193; MYOSINHEAVY.	
DR	ProDom; PD000355; myosin_head; 1.	
DR	SMART; SM00015; IQ; 1.	
DR	SMART; SM00242; MYSC; 1.	
DR	PROSITE; PS50096; IQ; 1.	
KW	Myosin; ATP-binding; Calmodulin-binding; Actin-binding;	
KW	Coiled coil; Alkylation; Multigene family; Disease mutation;	
KW	Deafness.	
FT	DOMAIN	1 778 MYOSIN HEAD-LIKE.
FT	DOMAIN	779 808 IQ.
FT	DOMAIN	837 1936 COILED COIL (POTENTIAL).
FT	NP_BIND	174 181 ATP (POTENTIAL).
FT	DOMAIN	654 676 ACTIN-BINDING.
FT	MOD_RES	694 694 ALKYLATION (SH-1) (POTENTIAL).
FT	MOD_RES	704 704 ALKYLATION (SH-2) (POTENTIAL).
FT	VARIANT	93 N -> K (IN MHA).
FT	VARIANT	702 702 /FTIG=VAR 010791.
FT	VARIANT	705 705 R -> C (IN FTNS).
FT	VARIANT	1155 1155 R -> H (IN DFNAL17).
FT	VARIANT	1165 1165 /FTIG=VAR 010792.
FT	VARIANT	1424 1424 /FTIG=VAR 010793.
FT	VARIANT	1841 1841 T -> I (IN MHA).
FT	VARIANT	53 55 EAI -> RGH (IN REF. 3).
FT	VARIANT	660 660 T -> S (IN REF. 3).
FT	VARIANT	869 869 T -> M (IN REF. 4).
FT	VARIANT	931 931 C -> Y (IN REF. 4).
FT	VARIANT	1240 1241 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010797.
FT	VARIANT	660 660 EAI -> RGH (IN REF. 3).
FT	VARIANT	869 869 T -> S (IN REF. 3).
FT	VARIANT	931 931 T -> M (IN REF. 4).
FT	VARIANT	1240 1241 C -> Y (IN REF. 4).
FT	VARIANT	1841 1841 KG -> GR (IN REF. 4).
FT	VARIANT	1424 1424 D -> H (IN FTNS).
FT	VARIANT	1841 1841 /FTIG=VAR 010796.
FT	VARIANT	53 55 E -> K (IN MHA).
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	VARIANT	1841 1841 C -> Y (IN REF. 4).
FT	VARIANT	1424 1424 KG -> GR (IN REF. 4).
FT	VARIANT	1841 1841 E -> K (IN MHA).
FT	VARIANT	53 55 /FTIG=VAR 010796.
FT	VARIANT	660 660 /FTIG=VAR 010797.
FT	VARIANT	869 869 EAI -> RGH (IN REF. 3).
FT	VARIANT	931 931 T -> S (IN REF. 3).
FT	VARIANT	1240 1241 T -> M (IN REF. 4).
FT	V	

```

DE Golgin-160 (Male-enhanced antigen-2) (MEA-2).
GN GOLGA3 OR MEA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Testis;
RX MEDLINE=97217683; PubMed=9063644;
RA Kondo M., Sutou S.;
RT "Cloning and molecular characterization of cDNA encoding a mouse
RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
RT autoantigen."
RL DNA Seq 7:71-82 (1997).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
CC DETECTABLE MALE ANTIGEN (SDM).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
CC LEYDIG CELLS, SPERMATOGENIA, OR SPERMATOCYTES.
CC -1- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-19 OR MET-30 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78270; BAA19612.1; --
DR MGD; MGI:96958; Golga3.
DR GO; GO:0005793; C:ER-Golgi intermediate compartment; IDA.
DR GO; GO:0000139; C:Golgi membrane; IDA.
DR GO; GO:0005515; P:protein binding activity; IPI.
DR GO; GO:0005515; P:developmental protein.
KW Spermatogenesis; Developmental protein.
FT DOMAIN 201 204 POLY-ALA.
SQ SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;

Query Match 16.7%; Score 78; DB 1; Length 1325;
Best Local Similarity 25.5%; Pred. No. 16;
Matches 28; Conservative 23; Mismatches 33; Indels 26; Gaps 4;

QY 4 ATLAQAGNFRISGDLKTDQVSTAG-----SLOQWRGAAG 43
DB 201 AARAHQDQNGEVRSDICSVSMESLAEFQDELLOILKDKRLEQVEALSJ 260

QY 44 TAAQAAVVRFOEANKQKDELDSINIRAGVOYGRADDEEQOALSOM 93
DB 261 EASQA----LQKAEALQAO-LAALSTRL-QAOVHSHSSQKQDSLSSEV 304

RESULT 9
K2M2 SHEEP STANDARD; PRT; 491 AA.
ID K2M2 SHEEP
AC P15241;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II microfibrillar, component 7C.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE
RX MEDLINE=90026244; PubMed=2803231;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;

```

```

RT "The amino acid sequence of component 7c, a type II intermediate-
RT filament protein from wool."
RL Biochem. J. 261:1015-1022(1989).
CC -1- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -1- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 109 HEAD.
FT DOMAIN 110 416 ROD.
FT DOMAIN 417 491 TAIL.
FT DOMAIN 110 144 COIL 1A.
FT DOMAIN 145 154 LINKER 1.
FT DOMAIN 155 255 COIL 1B.
FT DOMAIN 256 272 LINKER 12.
FT DOMAIN 273 416 COIL 2.
FT VARIANT 74 74 C -> G OR S.
FT VARIANT 80 80 C -> S.
FT VARIANT 144 144 F -> Y.
FT VARIANT 232 232 S -> V.
FT VARIANT 276 276 C -> D OR N.
FT VARIANT 284 284 Q -> H.
FT UNSURE 1 2 CG -> GC.
SQ SEQUENCE 491 AA; 53681 MW; A801771FE3831ABE CRC64;

Query Match 16.3%; Score 76; DB 1; Length 491;
Best Local Similarity 22.9%; Pred. No. 8.5;
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;

QY 12 NFERISGDLKTDQVSTAGSLQGWGRGAAGTAAQAAVVRFOEANKQKDELDSINI 71
DB 273 NDCIVAEIKAYQYDIIASRRAEAEWYRSKCEIKATVIRHGETLRTKEINELNEVI 332

QY 72 RQ--AGVOYSRADDEEQOALSQ 92
DB 333 QRLTAIVENAKQNSKLEAAVTQ 355

RESULT 10
TPR_HUMAN STANDARD; PRT; 2349 AA.
ID TPR_HUMAN
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain."
RN Oncogene 7:2329-2333(1992).
RL [2]
RP REVISIONS, AND CHARACTERIZATION.

```

RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526 (1994).
[3]
RN SEQUENCE OF 1-142 FROM N.A.
RP MEDLINE=88262257; PubMed=3387099;
RX King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RA "tpr homologues activate met and raf.";
RL Oncogene 2:617-619 (1988).
CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P2.
CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/TPRID282.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X66397; CAA47021.1; -
CC EMBL; Y00672; CAA68681.1; -
CC Genew; HGNC:12017; TPR.
CC MIM; 189940; -
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005643; C:nuclear pore; TAS.
CC GO; GO:0006606; P:protein-nucleus import; TAS.
CC Coiled coil; Proto-oncogene; Chromosomal translocation;
KW Nuclear protein; Transport.
FT DOMAIN 78 360
FT COILED COIL (POTENTIAL).
FT DOMAIN 422 571
FT COILED COIL (POTENTIAL).
FT DOMAIN 575 628
FT COILED COIL (POTENTIAL).
FT DOMAIN 758 805
FT COILED COIL (POTENTIAL).
FT DOMAIN 834 869
FT COILED COIL (POTENTIAL).
FT DOMAIN 934 979
FT COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064
FT COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166
FT COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241
FT COILED COIL (POTENTIAL).
FT DOMAIN 1282 1304
FT COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434
FT COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595
FT COILED COIL (POTENTIAL).
FT DOMAIN 527 530
FT POLY-SER.
FT DOMAIN 1833 1836
FT POLY-GLU.
FT DOMAIN 1957 1964
FT POLY-ASP.
FT DOMAIN 2235 2298
FT POLY-SER.
SQ SEQUENCE 2345 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 16.3%; Score 76; DB 1; Length 2349;
Best Local Similarity 26.2%; Pred. No. 45;
Matches 28; Conservative 16; Mismatches 39; Indels 24; Gaps 4;

QY 2 DAATLAQAGNPE-----RISG-----DLKTQIDQVESTAGSLQGWGAA 42
DB 900 EIALTKHLSNNEVQVASCSSQRTGKGFSNKKEDVDLVLSQLRTEEQVNDLKERLKTST 959
QY 43 GTAA--QAQVVRFOAANKQKQELDEISTNRQAGVQVSRADBEQQQ 87
DB 960 SNVEQYQAMVTSLESLNKKQVTEVRKNIE---VRLKESAEFTQ 1003

RESULT 11
FNR ANAVA STANDARD; PRT; 440 AA.
AC Q44549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin--NADP reductase (EC 1.18.1.2) (FNR).
GN PETH.
OS Anabaena variabilis.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
CX NCBI_TaxID=1172;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN-PC 7937 // ATCC 29413;
RA Mannan R.M., Matthijs H.C.P., Pakrasi H.B.;
RT "Cloning and molecular characterization of the peth gene in the
RT cyanobacterium Anabaena variabilis ATCC 29413.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
CC ferredoxin + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: May be bound to the thylakoid membrane or
CC anchored to the thylakoid-bound phycobilisomes.
CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
CC -!- SIMILARITY: Contains 1 cpcD-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L26346; AAA91046.1; -
CC HSP; P21890; IQUE.
CC InterPro; IPR001685; CpcD-like.
CC InterPro; IPR001834; Cyt B5 reductase.
CC InterPro; IPR003097; FAD binding.
CC InterPro; IPR001709; FPN_Cyt reductase.
CC InterPro; IPR001433; Oxred_FAD/NAD(P).
CC Pfam; PF01383; CpcD; 1.
CC Pfam; PF00667; FAD binding_1; 1.
CC Pfam; PF00175; NAD binding_1; 1.
CC PRINTS; PR00406; CYTBSRDTASE.
CC PRINTS; PR00371; FENCR.
CC ProDom; PD002828; CpcD-like_C; 1.
KW Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
KW Phycobilisome.
FT DOMAIN 1 98 CPCD-LIKE.
FT NP BIND 288 306 NADP (RIBOSE PART) (BY SIMILARITY).
SQ SEQUENCE 440 AA; 48826 MW; 660BAA2DCF59BB6 CRC64;

Query Match 16.2%; Score 75.5; DB 1; Length 440;
Best Local Similarity 30.0%; Pred. No. 8.4; 22; Indels 5; Gaps 1;
Matches 18; Conservative 15; Mismatches 15;

QY 33 SLOGQWRGAGTAQAQAVVRFOAANKQKQELDEISTNRQAG-----VOYSRADEBQQQ 87
DB 2 SNQGAFEAGANVESGRVFFVYVVGVRQNEETDQTNYPKRGSSVFIRVYPRNMQEMQR 61

RESULT 12
MYH9 RAT STANDARD; PRT; 1961 AA.
ID MYH9 RAT
AC Q62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,

```

DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
RP Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RL - FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPIN.
CC - SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC - DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC - SIMILARITY: Contains 1 myosin-like globular head domain.
CC - SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31463; AAA74950.1; -
CC HSPB; P10587; I3E2
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; Myosin_head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF00663; myosin_head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSc; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Alkylation; Multigene family.
CC DOMAIN 1 778 MYOSIN HEAD-LIKE.
CC FT DOMAIN 779 808
CC FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
CC FT NP_BIND 174 181 ATP (POTENTIAL).
CC FT DOMAIN 654 676 ACTIN-BINDING.
CC FT MOD_RES 594 594 ALKYLATION (SH-1) (POTENTIAL).
CC FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
CC SEQUENCE 1961 AA; 226336 MW; 9B9876D9681FB19E CRC64;

Query Match 16.2%; Score 75.5; DB 1; Length 1961;
Best Local Similarity 22.8%; Pred. No. 41;
Matches 26; Conservative 26; Mismatches 23; Indels 39; Gaps 4;

OY 14 ERISGDKLTQIDQVESTAGSLQGWRCAGTAQAQAVRFE-----AANKQ--- 60
DB 1785 ERQNKELKALQEMES---AVKYSK-ASTAALEAKIAQLEEQLDNETKERRQAASQVRR 1840
OY 61 -----KQELDEISTNIRQAGVYSRADEEQQALSQ 92
DB 1841 AEKLLKDVLLQVEDERNARQFKDQADKASTRLKQLKQLEEAEEAQAANASR 1894

RESULT 13
PYS1 ANASP
ID PYS1 ANASP
AC P07124;

```

```

DT 01-APR-1998 (Rel. 07, Created)
DT 01-APR-1998 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phycobilisome 8.9 kDa linker polypeptide, phycocyanin-associated, rod
DE (L-8.9/R) (Rod capping linker protein).
GN CPD OR ASR0531
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN SEQUENCE FROM N.A.
RA Belknap W.R., Haseikorn R.;
RP MEDLINE=87246520; PubMed=3109890;
RA Belknap W.R., Haseikorn R.;
RT "Cloning and light regulation of expression of the phycocyanin operon
RT of the cyanobacterium Anabaena.";
RL EMBO J. 6:871-884(1987).
RN SEQUENCE FROM N.A.
RP MEDLINE=21134097; PubMed=11237320;
RA Cai Y.A., Murphy J.T., Wedemayer G.J., Glazer A.N.;
RT "Recombinant phycobiliproteins. Recombinant C-phycocyanins equipped
RT with affinity tags, oligomerization, and biospecific recognition
RT domains.";
RL Anal. Biochem. 290:186-204 (2001).
RN SEQUENCE FROM N.A.
RP MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC - SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05239; CAA28865.1; -
CC EMBL; AF178757; AAG09319.1; -
CC EMBL; AP003592; BAB72489.1; -
CC FIR; AB1873; AB1873.
CC InterPro; IPR001685; CpcD-like.
CC Pfam; PF01383; CpcD; 1.
CC ProDom; PD002828; CpcD-like C; 1.
CC Phycobilisome; Photosynthesis; Complete proteome.
CC SEQUENCE 80 AA; 8895 MW; 52DF2D7DEF544D5 CRC64;

Query Match 16.0%; Score 74.5; DB 1; Length 80;
Best Local Similarity 35.6%; Pred. No. 1.7;
Matches 21; Conservative 10; Mismatches 21; Indels 7; Gaps 2;

OY 36 GQWRGAGTAQAQA--VVRFOEAANKQKQELDEISTNIRQAG-----VOXSRADDEQQQ 87
DB 3 QQTILGAGSVSSASRFRVEVGLRQSSSETDKNKYNIRSGSVFITVPYSRMNEEYQR 61

RESULT 14
FENR ANASO
ID FENR ANASO STANDARD; PRT; 440 AA.
AC P21890;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin-NADP reductase (EC 1.18.1.2) (FNR).
GN PETH.

```

OS Anabaena sp. (strain PCC 7119).
 OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 ON NCBI_TaxID=1169;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93344523; PubMed=8343609;
 RA Fillat M.F., Flores E., Gomez-Moreno C.;
 RT "Homology of the N-terminal domain of the peth gene product from
 RT Anabaena sp. PCC 7119 to the CpcD phycobilisome linker polypeptide.";
 RL Plant Mol. Biol. 22:725-729(1993).
 RN [2]
 RP SEQUENCE OF 137-440 FROM N.A.
 RX MEDLINE=91088322; PubMed=2124680;
 RA Fillat M.F., Bakker H.A.C., Welsbeek P.J.;
 RT "Sequence of the ferredoxin-NADP(+) reductase gene from Anabaena PCC
 RT 7119.";
 RL Nucleic Acids Res. 18:7161-7161(1990).
 RN [3]
 RP SEQUENCE OF 152-193.
 RX STRAIN=1403.46;
 RA MEDLINE=88132819; PubMed=3124746;
 RA Sancho J., Peleato M.L., Gomez-Moreno C., Edmondson D.E.;
 RT "Purification and properties of ferredoxin-NADP+ oxidoreductase from
 RT the nitrogen-fixing cyanobacteria Anabaena variabilis.";
 RL Arch. Biochem. Biophys. 260:200-207(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 137-440.
 RX MEDLINE=97045988; PubMed=8890910;
 RA Serre L., Vellieux F.M.D., Medina M., Gomez-Moreno C.,
 RA Fontecilla-Camps J.C., Frey M.;
 RT "X-ray structure of the ferredoxin:NADP+ reductase from the
 RT cyanobacterium Anabaena PCC 7119 at 1.8-A resolution, and
 RT crystallographic studies of NADP+ binding at 2.25-A resolution.";
 RL J. Mol. Biol. 263:20-39(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 146-440.
 RX MEDLINE=99119196; PubMed=9922134;
 RA Martinez-Julvez M., Hermoso J., Hurley J.K., Mayoral T.,
 RA Sanz-Aparicio J., Tollin G., Gomez-Moreno C., Medina M.;
 RT "Role of Arg100 and Arg264 from Anabaena PCC 7119 ferredoxin-NADP+
 RT reductase for optimal NADP+ binding and electron transfer.";
 RL Biochemistry 37:17680-17691(1998).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 137-440.
 RX MEDLINE=20114461; PubMed=10651039;
 RA Mayoral T., Medina M., Sanz-Aparicio J., Gomez-Moreno C.,
 RA Hermoso J.A.;
 RT "Structural basis of the catalytic role of Glu301 in Anabaena PCC
 RT 7119 ferredoxin-NADP+ reductase revealed by x-ray crystallography.";
 RL Proteins 38:60-69(2000).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.38 ANGSTROMS).
 RX MEDLINE=20508220; PubMed=11053838;
 RA Morales R., Kachalova G., Vellieux F., Charon M.-H., Frey M.;
 RT "Crystallographic studies of the interaction between the
 RT ferredoxin-NADP+ reductase and ferredoxin from the cyanobacterium
 RT Anabaena: looking for the elusive ferredoxin molecule.";
 RL Acta Crystallogr. D 56:1408-1412(2000).
 CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
 CC ferredoxin + NADPH.
 CC -!- COFACTOR: FAD.
 CC -!- SUBCELLULAR LOCATION: May be bound to the thylakoid membrane or
 CC anchored to the thylakoid-bound phycobilisomes.
 CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
 CC -!- SIMILARITY: Contains 1 cpcd-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X72394; CAA51088.1; -;
 DR EMBL; X54039; CAA37973.1; -;
 DR PDB; 1QUE; 15-MAY-97.
 DR PDB; 1QUF; 17-SEP-97.
 DR PDB; 1BJK; 18-NOV-98.
 DR PDB; 1B2R; 15-DEC-99.
 DR PDB; 1EWY; 07-FEB-01.
 DR PDB; 1BQE; 27-FEB-02.
 DR PDB; 1E62; 14-JUN-01.
 DR PDB; 1E63; 14-JUN-01.
 DR PDB; 1E64; 14-JUN-01.
 DR PDB; 1GJR; 27-JUN-02.
 DR PDB; 1GO2; 17-OCT-02.
 DR PDB; 1GR1; 24-OCT-02.
 DR PDB; 1H85; 28-NOV-01.
 DR PDB; 1QGH; 27-FEB-02.
 DR PDB; 1QHO; 27-FEB-02.
 DR InterPro; IPR001685; CpcD-like.
 DR InterPro; IPR001834; Cyt B5 reductase.
 DR InterPro; IPR001709; RPN_Cyt reductase.
 DR InterPro; IPR001433; Oxred_FAD/NAD(P).
 DR Pfam; PF01383; CpcD; 1.
 DR Pfam; PF00175; NAD_binding_1; 1.
 DR PRINTS; PR00406; CYTB5RDASE.
 DR PRINTS; PR00371; FPNCR.
 DR ProDom; PD002828; CpcD-like C; 1.
 DR Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
 KW Phycobilisome; 3D-structure. CPCD-LIKE.
 FT DOMAIN 1 80
 FT NP_BIND 288 306
 FT CONFLICT 180 180
 FT STRAND 152 153
 FT TURN 154 155
 FT STRAND 156 156
 FT STRAND 158 167
 FT TURN 171 172
 FT STRAND 177 183
 FT TURN 185 186
 FT TURN 193 194
 FT STRAND 196 200
 FT STRAND 204 204
 FT TURN 206 207
 FT STRAND 210 210
 FT STRAND 214 218
 FT TURN 222 227
 FT STRAND 231 237
 FT STRAND 240 241
 FT STRAND 250 251
 FT HELIX 253 259
 FT TURN 260 260
 FT TURN 263 264
 FT STRAND 266 273
 FT TURN 283 284
 FT STRAND 286 292
 FT HELIX 293 295
 FT HELIX 296 307
 FT HELIX 309 314
 FT TURN 316 317
 FT STRAND 323 330
 FT HELIX 333 335
 FT TURN 337 338
 FT HELIX 339 348
 FT TURN 350 352
 FT STRAND 353 359
 FT TURN 360 362
 FT STRAND 364 364
 FT TURN 366 367
 FT STRAND 370 370
 FT HELIX 373 379
 FT TURN 380 380
 FT HELIX 381 388
 FT TURN 389 389

FT TURN 391 392
FT TURN 393 399
FT TURN 401 402
FT TURN 404 416
FT TURN 417 419
FT TURN 422 431
FT TURN 432 433
FT TURN 435 440
SQ SEQUENCE 440 AA; 48865 MW; 8E1F61D0F09338B6 CRC64;
Query Match 15.8%; Score 73.5; DB 1; Length 440;
Best Local Similarity 30.0%; Pred. NO. 13;
Matches 18; Conservative 15; Mismatches 22; Indels 5; Gaps 1;
FT NP_BIND 288 306 NADP (RIBOSE PART) (BY SIMILARITY).
SQ SEQUENCE 440 AA; 48838 MW; BB07AA1B99295C98 CRC64;
Query Match 15.8%; Score 73.5; DB 1; Length 440;
Best Local Similarity 30.0%; Pred. NO. 13;
Matches 18; Conservative 15; Mismatches 22; Indels 5; Gaps 1;
QY 33 SLOGQWRGAAGTAAQAQAAVVFQEAANKQKQELDEISTNIRQAG-----VOYSRADEEQQQ 87
DB 2 SNQCAFDAANVESGSRVFVYEVVGMKQNEETDQTNPIRKSGSVFIRVPYINRMQEMQR 61
Search completed: November 21, 2003, 16:04:18
Job time : 2.12018 secs

QY 33 SLOGQWRGAAGTAAQAQAAVVFQEAANKQKQELDEISTNIRQAG-----VOYSRADEEQQQ 87
DB 2 SNQCAFDAANVESGSRVFVYEVVGMKQNEETDQTNPIRKSGSVFIRVPYINRMQEMQR 61

RESULT 15

FENR ANASP
ID FENR ANASP STANDARD; PRT; 440 AA.
AC P8658; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferredoxin--NADP reductase (EC 1.18.1.2) (FNR).
GN PETH OR ALL4121.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001)
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized
CC ferredoxin + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: May be bound to the thylakoid membrane or
CC anchored to the thylakoid-bound phycobilisomes.
CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
CC -!- SIMILARITY: Contains 1 cpd-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003595; BAB75820.1; -
CC PIR; AB2321; AB2321.
CC HSP; P21890; 1B2R.
CC InterPro; IPR001685; CpcD-like.
CC InterPro; IPR001834; Cyt B5 reductase.
CC InterPro; IPR001709; FEN_cyt reductse.
CC InterPro; IPR001433; Oxred_FAD/NAD(P).
CC Pfam; PF01383; CpcD; 1.
CC Pfam; PF00175; NAD_binding_1; 1.
CC PRINTS; PR00406; CYTB5RDTASE.
CC PRINTS; PR00371; FPNCR.
CC ProDom; PD002828; CpcD-like C; 1.
CC Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;
KW Phycobilisome; Complete proteome.
FT DOMAIN 1 80 CPCD-LIKE.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 4.6746 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-10

Perfect score: 466

Sequence: 1 TDAATLAQEAAGNFERISGDL.....VOYGRADDEQQQALLSQMGF 95

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriaph.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	24.5	106	16 Q8FS24	Q8FS24 corynebacte
2	93.5	20.1	1186	16 Q32BQ2	Q32BQ2 streptomyce
3	91	19.5	96	16 Q8G878	Q8G878 bifidobacte
4	91	19.5	108	16 Q8NST8	Q8NST8 corynebacte
5	90.5	19.4	410	2 Q9X367	Q9X367 bacillus an
6	90	19.3	750	3 Q42657	Q42657 schizosacch
7	87.5	18.8	97	16 Q33692	Q33692 mycobacteri
8	84.5	18.1	97	16 Q33267	Q33267 mycobacteri
9	84.5	18.1	102	16 Q8VJ87	Q8VJ87 mycobacteri
10	84	18.0	1200	11 Q921B9	Q921B9 mus musculu
11	83.5	17.9	188	5 Q77248	Q77248 spodioptera
12	81	17.4	914	16 Q8ED18	Q8ED18 shewanella
13	80.5	17.3	707	2 Q9F5N1	Q9F5N1 rhizobium m
14	80	17.2	528	5 Q26589	Q26589 schistosoma
15	80	17.2	1940	5 Q02456	Q02456 schistosoma
16	79.5	17.1	409	2 Q8VQM8	Q8VQM8 prevotellia

17	79.5	17.1	474	2	Q8GI70	Q8GI70 rhodopsendo
18	79.5	17.1	1999	11	Q63731	Q63731 rattus norv
19	79	17.0	508	6	Q28582	Q28582 ovis aries
20	78	16.7	178	5	O02415	O02415 agrius conv
21	78	16.7	1447	11	Q9QVT2	Q9QVT2 mus musculu
22	78	16.7	1487	11	Q9QVT3	Q9QVT3 mus musculu
23	77.5	16.6	790	2	Q8Z0F5	Q8Z0F5 salmonella
24	77	16.5	507	4	P78386	P78386 homo sapien
25	77	16.5	507	4	Q9NSB1	Q9NSB1 homo sapien
26	77	16.5	801	16	Q8P3Q3	Q8P3Q3 xanthomonas
27	76.5	16.4	459	16	Q8G7M4	Q8G7M4 bifidobacte
28	76.5	16.4	623	16	Q8D3R3	Q8D3R3 vibrio vuln
29	76.5	16.4	698	16	Q92RJ4	Q92RJ4 rhizobium m
30	76.5	16.4	1186	3	Q8NJJ2	Q8NJJ2 aspergillus
31	76	16.3	331	2	Q9XJH6	Q9XJH6 paracoccus
32	76	16.3	401	16	Q8FL30	Q8FL30 xanthomonas
33	76	16.3	671	13	Q9YHD4	Q9YHD4 rana catesb
34	76	16.3	1390	4	Q8WZA3	Q8WZA3 homo sapien
35	76	16.3	1498	4	Q8TDA9	Q8TDA9 homo sapien
36	76	16.3	1530	4	O43241	O43241 homo sapien
37	75.5	16.2	529	11	Q93TJ3	Q93TJ3 mus musculu
38	75.5	16.2	622	16	Q8D3R4	Q8D3R4 vibrio vuln
39	75.5	16.2	1960	11	Q8VDD5	Q8VDD5 mus musculu
40	75.5	16.2	2756	10	Q9LJ60	Q9LJ60 arabidopsis
41	75	16.1	163	16	Q9RY46	Q9RY46 deinococcus
42	75	16.1	507	11	Q9D7M4	Q9D7M4 mus musculu
43	75	16.1	507	11	Q922T6	Q922T6 mus musculu
44	75	16.1	507	11	Q8CE83	Q8CE83 mus musculu
45	75	16.1	1793	4	Q8TEP3	Q8TEP3 homo sapien

ALIGNMENTS

RESULT 1

Q8FS24	PRELIMINARY;	PRT;	106 AA.
ID	Q8FS24		
AC	Q8FS24;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Hypothetical protein.		
GN	CE0582.		
OS	Corynebacterium efficiens.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.		
OX	NCBI TaxID=152794;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Y8-314 / AJ 12310 / DSM 44549 / JCM 11189;		
RA	Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,		
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,		
RA	Usuda Y., Sugimoto S.		
RT	"The entire genomic sequence of Corynebacterium efficiens Y8-314.";		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AP005215; BAC17392.1; -		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 106 AA; 11382 MW; 73AF6CDBEA7838A6 CRC64;		

Query Match	24.5%;	Score 114;	DB 16;	Length 106;
Best Local Similarity	24.5%;	Pred. No. 0.0051;		
Matches	23;	Conservative	25;	Mismatches 46; Indels 0; Gaps 0;

Qy	1	TDAATLAQEAAGNTERISGDLKTDIDQVESTAGSQGWRGAAGTAQAQAAVVRFOEAAKQ 60
Db	7	TESDVMLATAGQVDDTNDQVQNELRLRGVVDVSRGWSAQQAQVDFALMNRWDSARQL 66

Qy	61	KOELDETSTNRQAGVQYSRADDEQQQALLSQMG 94
----	----	--------------------------------------

Db	67	QEALDSISTNIRANARSFDTSEADNAQAFSAVGG 100
----	----	--

RESULT 2

```
Q9ZBQ2
ID Q9ZBQ2 PRELIMINARY; PRT; 1186 AA.
AC Q9ZBQ2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative chromosome associated protein.
GN SC0577 OR SC7A1.21.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapeite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.; and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Chater K.F., Harris D.E., Cordero-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Chandra G., Chen C.W., Collins M.,
RA Harper D., Bateman A., Brown S., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939124; CAA22420.1; -.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR Complete proteome.
SQ SEQUENCE 1186 AA; 128723 MW; CB11027815373E99 CRC64;

Query Match 20.1%; Score 93.5; DB 16; Length 1186;
Best Local Similarity 33.0%; Pred. No. 5.1;
Matches 31; Conservative 15; Mismatches 41; Indels 7; Gaps 3;

QY 2 DAATLAEGNFERISGDLTKTQIDQVESTAGSLQGWGAGTA--AQAAVVFQEAANK 59
Db 691 ECAALVELGERRAAREKSSVAQ---QLGRLAGQARGAAGERSAALAEQAALDK 747

QY 60 QKQELDEISTNIRQAGVYSRADEEQQALSS 93
Db 748 ALTEVEELAE--RLVAEEMPVEEPTAARDRL 779

RESULT 3
Q8G878
ID Q8G878 PRELIMINARY; PRT; 96 AA.
AC Q8G878;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN BL0003.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwanen M.-C., Desiere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.;
RL "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
DR EMBL; AE014614; AAN23870.1; -.
DR Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
RW Hypothetical protein; Complete proteome
SQ SEQUENCE 96 AA; 10431 MW; 7D02B38C681F61C CRC64;

Query Match 19.5%; Score 91; DB 16; Length 96;
Best Local Similarity 27.3%; Pred. No. 0.53;
Matches 24; Conservative 17; Mismatches 37; Indels 10; Gaps 2;

QY 14 ERI---SDDLTKTQIDQVESTAG-----SLQGWGAGTAQAQAAVVFQEAANKQKE 63
Db 9 ERIQSSAAVATSIQIRQAVGVMTNLNALQDARWGSAAATQFTTAAVEQWRAQQQMEAS 68

QY 64 LDEISTNIRQAGVYSRADEEQQALSS 91
Db 69 LESIQRSITQASTVYADAEIQASRLFAS 96

RESULT 4
Q8N8T8
ID Q8N8T8 PRELIMINARY; PRT; 108 AA.
AC Q8N8T8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Cgl0579.
GN CGL0579.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005275; BAB97972.1; -.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 11644 MW; E0B3AFA03AE9D452 CRC64;

Query Match 19.5%; Score 91; DB 16; Length 108;
Best Local Similarity 19.8%; Pred. No. 0.6;
Matches 18; Conservative 26; Mismatches 47; Indels 0; Gaps 0;

QY 1 TDAATLAEGNFERISGDLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQ 60
Db 7 TESDYMLATASQVDDINDQVQGLSLRGVVDVSRGWSAGQAVSFDSLMMRWSSARQL 66

QY 61 KOELDEISTNIRQAGVYSRADEEQQALSS 91
Db 67 QEALASISDNIRHNARSFENTEADNSQAFNA 97
```


Query Match	Best Local Similarity	Score	DB 3	Length	750
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3					
QY 8 QEAG--NFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRQEAANKQKQELD 65					
DB 550 KOAGENHYNSLSXDYETQIKSLESL-----TNSQACVSPQEKINELNSQID 597					
QY 66 EISTNIROAGVOYRADEBQOQALS 90					
DB 598 ELKLNKANEKY-----QELAIS 616					
RESULT 7					
OS3692	PRELIMINARY;	PRT;	97 AA.		
AC OS3692;					
DT 01-JUN-1998 (TrEMBLrel. 06, Created)					
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)					
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)					
DE Hypothetical 9.8 kDa protein (PE family protein).					
GN RV0287 OR MT035.15 OR MT0300.					
OS Mycobacterium tuberculosis.					
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;					
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.					
OX NCBI_TaxID=1773;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=H37RV;					
RX MEDLINE=98295987; PubMed=9634230;					
RA Cole S.T., Bresch R., Parkhill J., Garnier T., Churcher C., Harris D.,					
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,					
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,					
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,					
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,					
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,					
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,					
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;					
RT "Deciphering the biology of Mycobacterium tuberculosis from the					
RL complete genome sequence.";					
RL Nature 393:537-544 (1998).					
RN [2]					
RP SEQUENCE FROM N.A.					
RC STRAIN=CDC 1551 / Oshkosh;					
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,					
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,					
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,					
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,					
RA Bishai W.;					
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and					
RL laboratory strains";					
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AL021930; CAA17362.1; -					
DR EMBL; AE006937; AAK44524.1; -					
DR TIGR; MT0300; -					
DR TubercuList; Rv0287; -					
KW Hypothetical protein; Complete proteome.					
SQ SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;					
Query Match	18.8%;	Score 87.5;	DB 16;	Length 97;	
Best Local Similarity	34.9%;	Pred. No. 1.1;			
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1					
QY 3 ATLAQAGNFRISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRQEAANKQK- 61					
DB 16 AKTVKNTSRSLVYHODYSQTEYIAS-----QWGSASSD-----RFVQMFNEAKP 61					
QY 62 -----QELDELSTNIROAGVOYRADE 83					
DB 62 MNFNILQELDKIAVELERAARVREADE 89					
Query Match	19.4%;	Score 90.5;	DB 2;	Length 410;	
Best Local Similarity	29.5%;	Pred. No. 2.9;			
Matches 26; Conservative 14; Mismatches 27; Indels 21; Gaps 3;					
QY 3 ATLAQAGNFRISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRQEAANKQK- 61					
DB 16 AKTVKNTSRSLVYHODYSQTEYIAS-----QWGSASSD-----RFVQMFNEAKP 61					
QY 62 -----QELDELSTNIROAGVOYRADE 83					
DB 62 MNFNILQELDKIAVELERAARVREADE 89					
Query Match	19.4%;	Score 90.5;	DB 2;	Length 410;	
Best Local Similarity	29.5%;	Pred. No. 2.9;			
Matches 26; Conservative 14; Mismatches 27; Indels 21; Gaps 3;					
QY 3 ATLAQAGNFRISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRQEAANKQK- 61					
DB 16 AKTVKNTSRSLVYHODYSQTEYIAS-----QWGSASSD-----RFVQMFNEAKP 61					
QY 62 -----QELDELSTNIROAGVOYRADE 83					

```

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 FAKAGLMRHTIGAEQAQNSAQAFHQGESSAAAFQAAHARFVAAAKVNTLLDVAQANLG 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSRADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 EAAGTYVAAD---AAAASYTGF 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
OS3267 PRELIMINARY; PRT; 97 AA.
ID OS3267;
AC OS3267;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PE-family protein.
GN RV3020C OR NTV012.34C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
DR EMBL; AL021287; CAAL16105.1;
DR TubercuList; RV3020C;
KW Complete proteome.
SQ SEQUENCE 97 AA; 9842 MW; BA9ECB3180EC17F2 CRC64;

Query Match 18.1%; Score 84.5; DB 16; Length 97;
Best Local Similarity 34.9%; Pred. No. 2.1;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 FAKAGLMRHTIGAEQAQNSAQAFHQGESSAAAFQAAHARFVAAAKVNTLLDVAQANLG 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSRADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 EAAGTYVAAD---AAAASYTGF 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
OS3267 PRELIMINARY; PRT; 102 AA.
ID OS3267;
AC OS3267;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PE family protein.
GN MT3105.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

```

```

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Onay M.L., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007129; AA47434.1; -.
DR TIGR; MT3105; -.
SQ SEQUENCE 102 AA; 10372 MW; 03CB984D6633D2E4 CRC64;

Query Match 18.1%; Score 84.5; DB 16; Length 102;
Best Local Similarity 34.9%; Pred. No. 2.2;
Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 13 PERISGLKTIQIDQVESTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNIR 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 FAKAGLMRHTIGAEQAQNSAQAFHQGESSAAAFQAAHARFVAAAKVNTLLDVAQANLG 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 QAGVQYSRADBEQQQALSSQMGF 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 EAAGTYVAAD---AAAASYTGF 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Q321B9 PRELIMINARY; PRT; 1200 AA.
ID Q321B9;
AC Q321B9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nuclear pore complex-associated protein Tpr (Fragment).
GN TPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=129/ola;
RA Sandblad L., Hunziker A., Cordes V.C.;
RT "Evolutionarily conserved mouse tpr is a single-copy gene located on
RT chromosome 1.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ298076; CAC40701.1; -.
DR InterPro; IPR005613; AIP3.
DR Pfam; PF03915; AIP3; 1.
FT NON_TER 1200 1200.
SQ SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;

Query Match 18.0%; Score 84; DB 11; Length 1200;
Best Local Similarity 27.1%; Pred. No. 37;
Matches 29; Conservative 16; Mismatches 38; Indels 24; Gaps 4;

QY 2 DAATLAQAGNFE-----RISG-----DLKTDQVESTAGSLQGWGAG 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 900 DIATLKQHLNNEAQLASQSTQRTKQGGFGDRDQVDDLLKSLQRAEEQVNDLKLKLTST 959
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 43 GTAQAQ--AAVVFQEAANKQKQELDEISTNIRQAGVQYSRADBEQQQ 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 960 SNVEQYRAMVTSLEDSLNKQVTEEVHKNIET---VRLKESAEFTQ 1003
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
O77248 PRELIMINARY; PRT; 188 AA.
ID O77248;
AC O77248;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Apolipoprotein-III.
GN APOLP-III.
OS Spodoptera litura.

```

```

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=69820;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99097579; PubMed=9880904;
RZ Kim E., Kim S.H., Choi C.S., Park Y.I., Kim H.R.;
RT "Cloning and expression of apolipophorin-III from the common cutworm,
RT Spodoptera litura";
RL Arch. Insect Biochem. Physiol. 39:166-173(1998).
DR EMBL; AF04582; AAC63377.1; -.
SQ SEQUENCE 188 AA; 20649 MW; 7A9AE36DF8E432C CRC64;

Query Match 17.9%; Score 83.5; DB 5; Length 188;
Best Local Similarity 26.7%; Pred. No. 5.3;
Matches 27; Conservative 20; Mismatches 45; Indels 9; Gaps 2;

QY 1 TDA-----ATLAQAGNFERISGDLKQIDQVESTAGSLQGWGAAGTAQAQAAVVFQEE 55
DB 88 TDANAKAKTALFQARQNLEKTAEDLRKSHPDVERQAGELRTKLQAAVQNTAQEVQKLAK 147
QY 56 AANKOKQELDEISTNRQAGVQVSRADDEQO-----QALSSQ 92
DB 148 VASVVEINEKLAPLKEAYENFSGHVEVQKKVHEALASKQ 188

RESULT 12
Q9ED18 PRELIMINARY; PRT; 914 AA.
AC Q9ED18
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Prophage lambdaSO, tail length tape measure protein.
GN H O S02953.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN
RP SEQUENCE FROM N.A.
RX STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim L.M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015734; AAN55966.1; -.
DR TIGR; S02953; -.
KW Complete proteome.
SQ SEQUENCE 914 AA; 98450 MW; 30189C1218EDCD9 CRC64;

Query Match 17.4%; Score 81; DB 16; Length 914;
Best Local Similarity 31.2%; Pred. No. 51;
Matches 20; Conservative 14; Mismatches 24; Indels 6; Gaps 1;

QY 24 IDQVESTAGSLQGWGAAGTAQAQAAVVFQEAANKQKQELDEISTNRQAGVQVSRAD 83
DB 11 LNWVAETGSFVEGQ-----TKAEALAKTEAAAKQKAEINRLGLDPLVAEYSKLDK 64
QY 84 EQQQ 87
DB 65 MEQQ 68

```

```

RESULT 13
Q9FSN1 PRELIMINARY; PRT; 707 AA.
AC Q9FSN1
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 23, Last annotation update)
DE Methyl-accepting chemotaxis protein McpU.
GN MCPU
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN
RP SEQUENCE FROM N.A.
RX Maschler P.F.J., Scharf B., Schmitt R.;
RT "Methyl-accepting chemotaxis protein McpU of Sinorhizobium meliloti";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF12876; AAG34154.1; -.
DR HSP; P02942; IQU7.
DR InterPro; IPR004010; Cache.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR003122; TarH.
DR Pfam; PF02743; Cache; 1.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF02203; TarH; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
SQ SEQUENCE 707 AA; 74391 MW; A82D7762FDB1FFCC CRC64;

Query Match 17.3%; Score 80.5; DB 2; Length 707;
Best Local Similarity 31.1%; Pred. No. 42;
Matches 33; Conservative 11; Mismatches 39; Indels 23; Gaps 4;

QY 4 ATLAQAGNFERISGDLKQIDQVESTAG-----SLOGWGAAGTAQAQAAVVFQEE 55
DB 570 AVVAQEVRELAQSAANAKEIKALITTSQGHSGVTLVGD----TCRALQAIWVEYQEI 625
QY 56 -----AANKQKQELDEIST--NIRQAGVQVSRADDEQOQALS 90
DB 626 NKHVSALVITATREQSTGLQEIINTAVMTDQGTQCNAAWVEQQTAS 671

RESULT 14
Q26589 PRELIMINARY; PRT; 528 AA.
AC Q26589
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Myosin II heavy chain (Fragment).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN
RP SEQUENCE FROM N.A.
RX STRAIN=Puerto Rican;
RX MEDLINE=93056536; PubMed=1431131;
RA Soisson L.M., Masterson C.P., Tom T.D., McNally M.T., Lowell G.H.,
RA Strand M.;
RT "Induction of protective immunity in mice using a 62-kDa recombinant
RT fragment of a Schistosoma mansoni surface antigen.";
RL J. Immunol. 149:3612-3620(1992).
DR EMBL; X65591; CAA46548.1; -.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 23.6728 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949

Sequence: 1 MYDFGALPPPEINSAHYAGP.....SCVLRVPRPYVMPHSPAG 391

Scoring table: BLOSUM62

Gapop:10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1949	100.0	391	18 AAW32449	Mycobacterium tube
2	1949	100.0	391	18 AAW32381	Mycobacterium tube
3	1949	100.0	391	19 AAW81702	M. tuberculosis im
4	1949	100.0	391	19 AAW64335	Mycobacterium tube
5	1949	100.0	391	20 AAY39132	M. tuberculosis an
6	1949	100.0	391	20 AAY39132	M. tuberculosis re
7	1949	100.0	391	20 AAY04778	Mycobacterium spec
8	1949	100.0	391	22 AAU01888	M. tuberculosis an
9	1949	100.0	391	23 AAE23707	Mycobacterium sp.

10	1949	100.0	391	23 AAE17571	Mycobacterium spec
11	1949	100.0	596	20 AAY32070	Mycobacterium tube
12	1949	100.0	596	23 AAE29710	Mycobacterium sp.
13	1949	100.0	596	23 AAE17574	Mycobacterium spec
14	1949	100.0	599	23 AAU74599	Antigenic fusion p
15	1949	100.0	600	20 AAY32068	Mycobacterium tube
16	1949	100.0	600	23 AAU74597	Antigenic fusion p
17	1949	100.0	729	22 AAO22142	Ral1-H9-32A fusion
18	1949	100.0	729	23 AAE29709	Mycobacterium sp.
19	1949	100.0	729	23 AAE17573	Mycobacterium spec
20	1949	100.0	744	22 AAU01902	M. tuberculosis an
21	1949	100.0	815	22 AAU01904	M. tuberculosis an
22	1949	100.0	930	23 AAE29731	Mycobacterium sp.
23	1945	99.8	788	22 AAU01903	M. tuberculosis an
24	1944	99.7	729	23 AAE04779	Mycobacterium spec
25	1944	99.7	729	23 AAE29708	Mycobacterium spec
26	1944	99.7	729	23 AAE17572	Mycobacterium spec
27	1931	99.1	729	20 AAY32059	Mycobacterium tube
28	1902.5	97.6	726	23 AAU74588	Antigenic fusion p
29	1652.5	84.8	396	19 AAW81704	M. tuberculosis im
30	1652.5	84.8	396	19 AAW64337	Mycobacterium tube
31	1652.5	84.8	396	20 AAY39134	M. tuberculosis an
32	1652.5	84.8	396	20 AAY38991	M. tuberculosis re
33	1486.5	76.3	359	19 AAW81703	M. tuberculosis im
34	1486.5	76.3	359	19 AAW64336	Mycobacterium tube
35	1486.5	76.3	359	20 AAY39133	M. tuberculosis an
36	1486.5	76.3	359	20 AAY38990	M. tuberculosis re
37	1474.5	75.7	341	18 AAW32451	Mycobacterium tube
38	1474.5	75.7	341	18 AAW32383	Mycobacterium tube
39	1187.5	60.9	358	23 AAU74591	Antigenic fusion p
40	1187	60.9	263	18 AAW32447	Mycobacterium tube
41	1187	60.9	263	18 AAW32379	Mycobacterium tube
42	1187	60.9	263	19 AAW81680	M. tuberculosis im
43	1187	60.9	263	19 AAW64317	Mycobacterium tube
44	1187	60.9	263	20 AAY32062	Mycobacterium tube
45	1187	60.9	263	20 AAY39121	M. tuberculosis an

ALIGNMENTS

RESULT 1
AAW32449

ID AAW32449 standard; Protein; 391 AA.

AC AAW32449;

XX 09-JAN-1998 (first entry)

DT Mycobacterium tuberculosis antigen TbH-9FL.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

XX WO9709428-A2.

PN 13-MAR-1997.

PD 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
XX Twardzik DR, Vedvick TH;

DR WPI; 1997-192903/17.
 DR N-PSDB; AAT91521.
 XX
 PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also
 PT for diagnosis
 XX
 PS Example 3; Page 138-139; 168pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tbh-9FL The immunogenic protein, and fusion proteins
 CC containing one or more of the proteins or one of the proteins plus
 CC EGAT-6, are useful in vaccines, preferably when formulated with a
 CC non-specific adjuvant, to induce an immune response against
 CC M.tuberculosis (for treatment or prevention).
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 18; Length 391;
 Best Local Similarity 100.0%; Pred. No. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Qy 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Qy 121 LIATNLGQNTPAIVNAEYEGEMWADAAAMFCYAAATATATATLLPFEAPEMTSAGG 180
 Db 121 LIATNLGQNTPAIVNAEYEGEMWADAAAMFCYAAATATATATLLPFEAPEMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGKWKTVSPHRSPI 240
 Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGKWKTVSPHRSPI 240
 Qy 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Qy 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Db 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Qy 361 GQMGARAGGGLSGVLRVPPPYVMPHSPAAG 391
 Db 361 GQMGARAGGGLSGVLRVPPPYVMPHSPAAG 391
 RESULT 2
 AAW32381
 ID AAW32381 standard; Protein; 391 AA.
 XX
 AC AAW32381;
 XX
 DT 13-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbh-9FL.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.
 CS Mycobacterium tuberculosis.
 XX
 FN WO9709429-A2.
 XX
 PD 13-MAR-1997.
 XX

PF 30-AUG-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 CC Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 DR WPI; 1997-192904/17.
 DR N-PSDB; AAT91455.
 XX
 CC New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 XX Example 3; Page 150-152; 190pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis
 CC antigen, Tbh-9FL The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic
 CC polypeptide can be used as diagnostic primers or probes and agents
 CC that bind to the antigen, especially monoclonal antibodies or
 CC equivalent polyclonal antibodies, are also used for diagnosis.
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 18; Length 391;
 Best Local Similarity 100.0%; Pred. No. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 Qy 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Db 61 SSAGLMVAASPPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 Qy 121 LIATNLGQNTPAIVNAEYEGEMWADAAAMFCYAAATATATATLLPFEAPEMTSAGG 180
 Db 121 LIATNLGQNTPAIVNAEYEGEMWADAAAMFCYAAATATATATLLPFEAPEMTSAGG 180
 Qy 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGKWKTVSPHRSPI 240
 Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGKWKTVSPHRSPI 240
 Qy 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Db 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSSLGSSG 300
 Qy 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Db 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQMLGGLPV 360
 Qy 361 GQMGARAGGGLSGVLRVPPPYVMPHSPAAG 391
 Db 361 GQMGARAGGGLSGVLRVPPPYVMPHSPAAG 391
 RESULT 3
 AAW81702
 ID AAW81702 standard; Protein; 391 AA.
 XX
 AC AAW81702;
 XX

```

27-JAN-1999 (first entry)
XX DE M. tuberculosis immunogenic polypeptide TbH-9FL.
XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX KW vaccine; pharmaceutical; infection; diagnosis.
XX OS Mycobacterium tuberculosis.
XX PN WC9816646-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18293.
XX PR 13-MAR-1997; 97US-0818112.
XX PR 11-OCT-1996; 96US-0730510.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-261042/23.
XX DR N-PSDB; AAV64503.
XX CC Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and for diagnosis, treatment and prevention of tuberculosis
XX PS Example 3B; Page 128-129; 230pp; English.
XX CC This sequence represents an immunogenic portion of a soluble
XX CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX CC for inducing protective immunity against tuberculosis (TB). This
XX CC sequence can be formulated into vaccines and/or pharmaceutical
XX CC compositions for immunising against M. tuberculosis infection or may
XX CC be used for the diagnosis of tuberculosis.
XX SQ Sequence 391 AA;
    Query Match 100.0%; Score 1949; DB 19; Length 391;
    Best Local Similarity 100.0%; Pred. No. 4.8e-143;
    Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAAAAPYVAMSVTAQQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAMSVTAQQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPENTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPENTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMWSLGSLSGSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMWSLGSLSGSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAAEERGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAAEERGQMLGGLPV 360
QY 361 GQMGARAGGLGVLVRVPPRVPVMPHSPAAG 391
DB 361 GQMGARAGGLGVLVRVPPRVPVMPHSPAAG 391

RESULT 4
AAW64335
ID AAW64335 standard; Protein; 391 AA.
XX AC AAW64335;
XX DT 09-NOV-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen TbH-9FL.
XX KW Tuberculosis; infection; diagnosis; antigen; TbH-9FL.
XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44395.
XX CC New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX PS Example 3; Page 133-135; 250pp; English.
XX CC This polypeptide comprises Mycobacterium tuberculosis antigen
XX CC TbH-9FL. It is encoded by genomic DNA (see AAV44395) isolated from a
XX CC M. tuberculosis strain H37Rv genomic library using a probe from
XX CC clone TbH-9 (see AAV44371). The invention relates to compositions
XX CC and methods for diagnosing tuberculosis. It provides polypeptides
XX CC (see AAW64291-W64379) comprising an antigenic portion of a soluble
XX CC M. tuberculosis antigen, or an immunogenic portion of an M.
XX CC tuberculosis antigen, as well as DNA sequences encoding such
XX CC polypeptides, recombinant expression vectors and transformed or
XX CC transfected host cells. Also claimed are methods and diagnostic
XX CC kits for detecting M. tuberculosis infection in a patient using
XX CC these polypeptides, antibodies or oligonucleotide probes and
XX CC primers, for the diagnosis of tuberculosis.
XX SQ Sequence 391 AA;
    Query Match 100.0%; Score 1949; DB 19; Length 391;
    Best Local Similarity 100.0%; Pred. No. 4.8e-143;
    Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAAAAPYVAMSVTAQQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAMSVTAQQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPENTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPENTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
```

QY 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
 DB 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 DB 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 DB 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 RESULT 5
 AAY39132
 ID AAY39132 standard; Protein; 391 AA.
 XX
 AC AAY39132;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen TbH-9FL amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI; 1999-527409/44.
 DR N-PSDB; AAZ19305.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 XX skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Example 3; Page 123-124; 299pp; English.
 CC
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 20; Length 391;
 Best Local Similarity 100.0%; Pred. NO. 4.8e-143;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDGALPPEINARMYAGPQSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 DB 1 MVDGALPPEINARMYAGPQSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60

QY 51 SSAGLWVAASPYVAMSVTAGQAEHTAAQVRVAAAAVETAYGLTVPPVIAENRAELMI 120
 DB 51 SSAGLWVAASPYVAMSVTAGQAEHTAAQVRVAAAAVETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGONTFPAJAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 DB 121 LIATNLLGONTFPAJAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSFIN 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSFIN 240
 QY 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 DB 241 MVSNNHMSMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 DB 301 LGGVAAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAERGPQGMGLGLPV 360
 QY 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 DB 361 GOMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; Protein; 391 AA.
 XX
 AC AAY38989;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein TbH-9FL.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO9942118-A2.
 PN
 XX 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-US03285.
 PF
 XX 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 XX WPI; 1999-527416/44.
 DR N-PSDB; AAZ19093.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 3; Page 168-169; 323pp; English.
 CC
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 20; Length 391;

PS Example 2; Page 151-152; 168pp; English.

XX The sequence represents Mycobacterium tuberculosis TBH9 (also known

CC as Mtb39A), an M. tuberculosis antigen. Compositions comprising at least

CC 2 heterologous antigens, as a fusion protein, and vectors expressing the

CC fusion proteins are used as vaccines to prophylactically immunise

CC mammals (especially humans) against infection by Mycobacteria. The

CC compositions contain at least 2 heterologous antigens that increase the

CC serological sensitivity of individuals infected with tuberculosis, a

CC disease frequently affecting patients with acquired immunodeficiency

CC disease, AIDS.

XX Sequence 391 AA;

SQ Query Match 100.0%; Score 1949; DB 22; Length 391;

Best Local Similarity 100.0%; Pred. No. 4.8e-143;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASATQSVVWGLTVGSWIG 60

DB 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASATQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPPYVAMNSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

DB 61 SSAGLMVAASPPYVAMNSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240

DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSG 300

DB 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

QY 361 GQMGARAGGSLGVLVPPPPYVMPHSPAAG 391

DB 361 GQMGARAGGSLGVLVPPPPYVMPHSPAAG 391

RESULT 9

AAE29707

ID AAE29707 standard; Protein; 391 AA.

AC AAE29707;

XX 27-JAN-2003. (first entry)

DT Mycobacterium sp. TBH9FL antigenic protein.

DE Vaccine; immunity; diagnostic agent; gene therapy; TBH9FL antigen.

KW Mycobacterium sp.

XX Mycobacterium sp.

OS WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US08223.

XX 13-MAR-2001; 2001US-275837P.

PR (CORI-) CORIXA CORP.

XX Skeiky Y, Brannon M, Guderian J;

DR WPI; 2002-759844/82.

DR N-PSDB; AAD47082.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA,

PT LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective

PT immunity against pathogenic microorganisms e.g. Leishmania and

XX Mycobacterium tuberculosis

PS Disclosure; Page 86-87; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a

CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous

CC polynucleotide sequence encoding an antigen or an antigenic fragment from

CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a

CC polypeptide or its fragment. The Leishmania polynucleotide is selected

CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are

CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion

CC polypeptides are used for enhancing the expression of polynucleotides,

CC as in vivo diagnostic agents and for raising antibodies in a non-human

CC animal. The invention is used in gene therapy. The present sequence is

CC Mycobacterium sp. TBH9FL antigenic protein.

XX Sequence 391 AA;

SQ Query Match 100.0%; Score 1949; DB 23; Length 391;

Best Local Similarity 100.0%; Pred. No. 4.8e-143;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASATQSVVWGLTVGSWIG 60

DB 1 MVDGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASATQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPPYVAMNSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

DB 61 SSAGLMVAASPPYVAMNSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240

DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQLOAQTOGTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSG 300

DB 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

QY 361 GQMGARAGGSLGVLVPPPPYVMPHSPAAG 391

DB 361 GQMGARAGGSLGVLVPPPPYVMPHSPAAG 391

RESULT 10

AAE17571

ID AAE17571 standard; Protein; 391 AA.

AC AAE17571;

XX 22-APR-2002 (first entry)

DT Mycobacterium species MTB39 (TBH9) protein #2.

DE Fusion protein; antigen; serological sensitivity; immune response;

XX tuberculosis; infection; vaccine; MTB39; TBH9 protein.

XX Mycobacterium sp.

OS

XX W0200198460-A2.
XX 27-DEC-2001.
XX 20-JUN-2001; 2001WO-US19959.
XX 20-JUN-2000; 2000US-0597796.
XX 01-FEB-2001; 2001US-2657372.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
XX N-PSDB; RAD28341.
XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a
XX subject -
XX Claim 83; Page 102-103; 136pp; English.
XX The present invention relates to fusion proteins containing at least
XX two Mycobacterium species antigens, nucleotides encoding them and
XX compositions comprising such fusion proteins. The present invention
XX particularly relates to nucleic acids encoding fusion proteins that
XX include two or more individual M. tuberculosis antigens which increase
XX the serological sensitivity of sera from individuals infected with
XX tuberculosis and methods for their use in diagnosis, prevention and
XX treatment of tuberculosis infection. Sequences of the invention are
XX useful for eliciting an immune response in a mammal, e.g., human,
XX immunised with BCG. They are useful in the diagnosis, treatment and
XX prevention of Mycobacterium infection. The fusion proteins and the
XX polynucleotides are useful as diagnostic tools in patients infected
XX with Mycobacterium, in vitro and in vivo assays for detecting humoral
XX antibodies or cell-mediated immunity against M. tuberculosis, for the
XX diagnosis of an infection or monitoring of disease progression, as
XX immunogens to generate or elicit a protective immune response in a
XX patient and for raising anti-M. tuberculosis antibodies in a non-human
XX animal. Sequences of the invention are also used as vaccines. MTB32A
XX fusion proteins of the invention are useful as in vivo diagnostic agents
XX for intradermal skin test. The present sequence is Mycobacterium species
XX MTB39 (TbH9) protein.
XX SQ Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 23; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.8e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDPGALPPPEINSGARMYAGPGSASLVAAQWMDVSASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDPGALPPPEINSGARMYAGPGSASLVAAQWMDVSASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLWVAASBPVWMSVTAQAGELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLWVAASBPVWMSVTAQAGELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
QY 121 LIATNLGQNTPALVAVNEAYEGWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLGQNTPALVAVNEAYEGWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQLMNVPAQLOLAQFTGTTTSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVEASDTAAANQLMNVPAQLOLAQFTGTTTSSKLGGLWKTVPSPHSPISN 240
QY 241 MVSNNHMTKNSGVSMNTTLLSSMLKGFAPAAAQAVTAAGNGVRAMSSLSGSSG 300
DB 241 MVSNNHMTKNSGVSMNTTLLSSMLKGFAPAAAQAVTAAGNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTFAARALPLTSLTSAERGPQMLGGLPV 360

DB 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTFAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
RESULT 11
AAY32070
ID AAY32070 standard; Protein; 596 AA.
XX
XX AAY32070;
XX
XX 17-JAN-2000 (first entry)
XX
XX Mycobacterium tuberculosis antigen fusion protein Mtb59f.
XX Tuberculosis; antigen; fusion protein; Mtb59f; TbH9; Ra35;
XX diagnosis; therapy; vaccine; immunogen.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..8
XX FT /note= "Met/His tag"
XX FT Protein 9..140
XX FT /note= "Ra12"
XX FT Protein 143..596
XX FT /note= "TbH9"
XX
XX W09951748-A2.
XX
XX 14-OCT-1999.
XX
XX 07-APR-1999; 99WO-US07717.
XX
XX 07-APR-1999; 98US-0056556.
XX 30-DEC-1999; 98US-0223040.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Alderson M, Campos-Neto A;
XX
XX WPI; 1999-601610/51.
XX N-PSDB; AA220205.
XX
XX New fusion proteins useful for diagnosis, prevention and treatment of
XX tuberculosis -
XX Claim 1; Fig 12A-B; 83pp; English.
XX
XX This sequence represents a recombinant Mycobacterium tuberculosis
XX bi-antigen fusion protein, termed Mtb59f, composed of the antigens
XX TbH9 and Ra35. The fusion protein is expressed in host cells
XX using a vector carrying a polynucleotide (see AAZ20205) comprising
XX the coding sequences for the 2 antigens. The invention provides
XX fusion proteins (see AAY32059-71) containing at least 2 M.
XX tuberculosis antigens. The new fusion proteins and polynucleotides
XX encoding them are useful as vaccines for preventing tuberculosis
XX (claimed), for diagnosis (via in vitro assays or intradermal skin
XX tests for detection of anti-M. tuberculosis antibodies), monitoring
XX of disease progression, and treatment of tuberculosis. They are
XX more effective immunogens than mixtures of the individual protein
XX components.
XX
XX Sequence 596 AA;
Query Match 100.0%; Score 1949; DB 20; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSGARMYAGPGSASLVAAQWMDVSASDLFSAASAFQSVVWGLTVGSWIG 60

Db 9 MVDFGALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAAASPYVAAWMSVTAGQAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLMVAAAASPYVAAWMSVTAGQAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATATLLPFEAPEMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATATLLPFEAPEMTSAGG 188
 QY 181 LIEQAAAEEASDTAAANQLMNNVPAALQQLAQTGTTSSKLGGLWKTVPSPHSPISN 240
 Db 189 LIEQAAAEEASDTAAANQLMNNVPAALQQLAQTGTTSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 300
 Db 249 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 308
 QY 301 LGGVAAANLGRAASVGSLSVPOAANAQAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAANAQAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 12

AAE29710
 ID AAE29710 standard; Protein; 596 AA.

XX AC AAE29710;

XX DT 27-JAN-2003 (first entry)

XX DE Mycobacterium sp. MTB59F fusion protein.

XX KW Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen;
 KW Ra35; MTB59F; fusion protein.

XX OS Chimeric - Mycobacterium sp.

XX OS Chimeric - Mycobacterium tuberculosis.

XX PN WO200272792-A2.

XX PD 19-SEP-2002.

XX PF 13-MAR-2002; 2002WO-US08223.

XX PR 13-MAR-2001; 2001US-275837P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Brannon M, Guderian J;

XX PR WPI; 2002-759844/82.

XX DR N-PSDB; AAD47086.

XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA.
 PT LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
 PT immunity against pathogenic microorganisms e.g. Leishmania and
 PT Mycobacterium tuberculosis

XX PS Disclosure; Page 98-99; 155pp; English.

XX CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides,
 CC as in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB59F fusion protein. This fusion protein comprises Ra35 protein from
 CC Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp.

XX SQ Sequence 596 AA;

Query Match 100.0%; Score 1949; DB 23; Length 596;

Best Local Similarity 100.0%; Pred. No. 8.1e-143; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDFGALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 Db 9 MVDFGALPPEINARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAAASPYVAAWMSVTAGQAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLMVAAAASPYVAAWMSVTAGQAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLLPFEAPEMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLLPFEAPEMTSAGG 188
 QY 181 LIEQAAAEEASDTAAANQLMNNVPAALQQLAQTGTTSSKLGGLWKTVPSPHSPISN 240
 Db 189 LIEQAAAEEASDTAAANQLMNNVPAALQQLAQTGTTSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 300
 Db 249 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 308
 QY 301 LGGVAAANLGRAASVGSLSVPOAANAQAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAANAQAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 13

AAE17574
 ID AAE17574 standard; Protein; 596 AA.

XX AC AAE17574;

XX DT 22-APR-2002 (first entry)

XX DE Mycobacterium species MTB59F fusion protein.

XX KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB59F; TBH9-Ra35 protein.

XX OS Mycobacterium sp.

XX PN WO200198460-A2.

XX PD 27-DEC-2001.

XX PF 20-JUN-2001; 2001WO-US19959.

XX PR 20-JUN-2000; 2000US-0597796.

XX PR 01-FEB-2001; 2001US-265737P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Reed S, Alderson M;

XX DR WPI; 2002-147798/19.

XX DR N-PSDB; AAD28344.

PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject -
XX
XX
PS Claim 5; Page 114-115; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB59F (TbH9-Ra35) fusion protein.
XX
XX
SQ Sequence 596 AA;
Query Match 100.0%; Score 1949; DB 23; Length 596;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 9 MVDFGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGLMVAASPYVAVMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
DB 69 SSAGLMVAASPYVAVMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
QY 241 MVSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 300
DB 249 MVSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 308
QY 301 LGGVAAANLGRAASVGLSVPOAWAANQAVTPAARALPLTSLTSAARPGQWGLGGLPV 360
DB 309 LGGVAAANLGRAASVGLSVPOAWAANQAVTPAARALPLTSLTSAARPGQWGLGGLPV 368
QY 361 GQMGARAGGGLSVGLRVPPRPYVPHSPAAG 391
DB 369 GQMGARAGGGLSVGLRVPPRPYVPHSPAAG 399
RESULT 14
AAU74599
ID AAU74599 standard; Protein; 599 AA.
XX
XX AC AAU74599;
XX DT
XX DT 08-MAY-2002 (first entry)
XX DE Antigenic fusion protein Tb59-Ra35 (Mtb59f).
XX

KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculosatic; immunogen; vaccine; Tb59-Ra35; Mtb59f.
XX
XX
OS Chimeric - Mycobacterium tuberculosis.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 597
FT /label= OTHER
FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
XX
XX
FN US2002009459-A1.
XX
PD 24-JAN-2002.
XX
PF 07-APR-1999; 99US-0287849.
XX
PR 13-MAR-1997; 97US-0818112.
PR 01-OCT-1997; 97US-0942578.
PR 18-FEB-1998; 98US-0025197.
PR 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX
XX (REED/) REED S G.
PA (SKEI/) SKEIKY Y A.
PA (DILL/) DILLON D C.
PA (ALDE/) ALDERSON M.
PA (CAMF/) CAMPOS-NETO A.
XX
PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
XX
XX WPI; 2002-171134/22.
DR N-PSDB; ABK14139.
XX
XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
PT diagnosing, treating or preventing M. tuberculosis infection,
PT particularly as vaccine for treating or preventing tuberculosis
XX
PS Claim 1; Fig 12; 62pp; English.
XX
XX The invention relates to a purified polypeptide which induces an immune
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
CC useful for diagnosing, treating or preventing M. tuberculosis infection,
CC particularly tuberculosis infection. In particular, the polypeptides are
CC useful as a vaccine formulation with an adjuvant to afford long-term
CC protection in animals against the development of tuberculosis. The
CC protein coding sequence may be used to encode a protein product for use
CC as an immunogen to induce and/or enhance an immune response to M.
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
CC of the invention.
XX
SQ Sequence 599 AA;
Query Match 100.0%; Score 1949; DB 23; Length 599;
Best Local Similarity 100.0%; Pred. No. 8.1e-143;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 9 MVDFGALPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGLMVAASPYVAVMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
DB 69 SSAGLMVAASPYVAVMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
QY 241 MVSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 300

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 8.24558 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949

Sequence: 1 MVDFGALPEINARMVAGP.....SGVLVPPRPVYMPHSPAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-596-102
5	1944	99.7	729	4	US-09-223-040-2
6	1652.5	84.8	396	3	US-08-818-112-111
7	1652.5	84.8	396	4	US-08-818-111-106
8	1652.5	84.8	396	4	US-09-056-556-111
9	1652.5	84.8	396	4	US-09-072-596-106
10	1486.5	76.3	359	3	US-08-818-112-109
11	1486.5	76.3	359	4	US-08-818-111-104
12	1486.5	76.3	359	4	US-09-056-556-109
13	1486.5	76.3	359	4	US-09-072-596-104
14	1187	60.9	263	3	US-08-818-112-91
15	1187	60.9	263	4	US-08-818-111-92
16	1187	60.9	263	4	US-09-056-556-91
17	1187	60.9	263	4	US-09-072-596-92
18	766.5	39.3	263	4	US-09-073-009-126
19	603	30.9	423	4	US-09-073-009-142
20	424.5	21.8	943	4	US-09-477-135A-131
21	424	21.8	141	4	US-09-073-009-15
22	381.5	19.6	204	4	US-08-311-731A-57
23	377.5	19.4	208	4	US-08-311-731A-208
24	314	16.1	368	3	US-08-818-112-114
25	314	16.1	368	4	US-08-818-111-109
26	314	16.1	368	4	US-09-056-556-114
27	314	16.1	368	4	US-09-072-596-109

28 261 13.4 1271 1 US-08-095-734-2 Sequence 2, Appli
29 261 13.4 1271 2 US-08-444-623-2 Sequence 2, Appli
30 261 13.4 1271 3 US-08-471-869-2 Sequence 2, Appli
31 261 13.4 1271 4 US-09-342-563-2 Sequence 2, Appli
32 261 13.4 1271 5 PCT-US94-08267-2 Sequence 2, Appli
33 215 11.0 352 4 US-09-073-009-14 Sequence 14, Appli
34 186.5 9.6 943 4 US-09-056-556-204 Sequence 104, App
35 186.5 9.6 943 4 US-09-072-596-199 Sequence 199, App
36 145 7.4 800 4 US-09-252-991A-20437 Sequence 20437, A
37 143.5 7.2 228 4 US-09-477-135A-128 Sequence 128, App
38 141 7.2 738 3 US-08-864-038A-3 Sequence 3, Appli
39 139.5 6.7 792 2 US-08-678-039A-40 Sequence 40, Appli
40 131.5 6.7 826 4 US-09-894-998A-47 Sequence 47, Appli
41 129.5 6.6 1186 2 US-08-861-464-8 Sequence 8, Appli
42 129.5 6.6 1186 2 US-08-396-001-8 Sequence 8, Appli
43 129.5 6.6 1186 3 US-09-323-433A-9 Sequence 8, Appli
44 128 6.6 731 4 US-09-340-736E-1 Sequence 1, Appli
45 126 6.5 731 2 US-08-911-364-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Sxelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred.No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSG 300
Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOQAAWAAANQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOQAAWAAANQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLVPRPVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLVPRPVMPHSPAAG 391

RESULT 2

US-08-818-111-102
; Sequence 102, Application US/0818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MYDFGALPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAOPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSG 300
Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOQAAWAAANQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOQAAWAAANQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLVPRPVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLVPRPVMPHSPAAG 391

RESULT 3

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-056-556-107

Query Match	100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity	100.0%; Pred. No. 2.7e-154;
Matches 391; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
1	MVDFCALPEINSAARMYAGPSASIVAAQWSDSVASDLFSAAGAFQSVVWGLTVGSWIG 60
1	MVDFCALPEINSAARMYAGPSASIVAAQWSDSVASDLFSAAGAFQSVVWGLTVGSWIG 60
61	SSAGLWMAAASPYVAWMSVTAGQABELTAAQVRVAAAAAYETAYGLTVPPIVAENRAELMI 120
61	SSAGLWMAAASPYVAWMSVTAGQABELTAAQVRVAAAAAYETAYGLTVPPIVAENRAELMI 120
121	LIATNLGQNTPAIAVNAEAYGEMWAQDAAMFGEYARATATATATLLEPEAPEMTSAGG 180
121	LIATNLGQNTPAIAVNAEAYGEMWAQDAAMFGEYARATATATATLLEPEAPEMTSAGG 180
181	LLEQAAAYVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSVPHRSPISN 240
181	LLEQAAAYVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSVPHRSPISN 240
241	MVSMANNHSMVTNSGVSMNTILSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSG 300
241	MVSMANNHSMVTNSGVSMNTILSSMLKGFAPAAAAQAVQTAAQNGVRAMSSLGSSLGSSG 300
301	LGGGVAAALNGLRAASVGSLSVPOAWAAAQNAVTPAARALPLTSLTSAAGRGQMLGGPLPV 360
301	LGGGVAAALNGLRAASVGSLSVPOAWAAAQNAVTPAARALPLTSLTSAAGRGQMLGGPLPV 360
361	GOMGARAGGGLSGVLVRPPRPVYMPHSPAAAG 391
361	GOMGARAGGGLSGVLVRPPRPVYMPHSPAAAG 391

RESULT 4

US-09-072-596-102
Sequence 102, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOCS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

```

, TELEFAX: (206) 682-6031
, INFORMATION FOR SEQ ID NO: 102:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 391 amino acids
, TYPE: amino acid
, STRANDEDNESS: single
, TOPOLOGY: linear
US-09-072-596-102

```

Query Match	100.0%;	Score 1949;	DB 4;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 2.7e-154;		
Matches 391;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MVDFGALPPINISARMYAGFSGASLVAAQAQWSDLSAASAFOSVVMGLTVGSWIG	60	
Db	1	MVDFGALPPINISARMYAGFSGASLVAAQAQWSDLSAASAFOSVVMGLTVGSWIG	60	
Qy	61	SSAGLMVAAASPYYVAMNSVTTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	120	
Db	61	SSAGLMVAAASPYYVAMNSVTTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	120	
Qy	121	LIATNLGQONTPALIYNVEAEYGEWGAQDAAMCYRAATATATATLPPFEAPEMTSAGG	180	
Db	121	LIATNLGQONTPALIYNVEAEYGEWGAQDAAMCYRAATATATATLPPFEAPEMTSAGG	180	
Qy	181	LLECAAAVEASDITAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISN	240	
Db	181	LLECAAAVEASDITAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISN	240	
Qy	241	MVSMANNHMSMTNSGVSMNTVLSMLKGFAPAAAAQAVQTAAQNGVRAMSSLSGSLGSSG	300	
Db	241	MVSMANNHMSMTNSGVSMNTVLSMLKGFAPAAAAQAVQTAAQNGVRAMSSLSGSLGSSG	300	
Qy	301	LGCGVAANLGRASVSGLSVPQAWAANQAVTPAARALPLTSLTSAERGPCQMLGGLPVP	360	
Db	301	LGCGVAANLGRASVSGLSVPQAWAANQAVTPAARALPLTSLTSAERGPCQMLGGLPVP	360	
Qy	361	GQMGARAGGSLGVLVRPPRPYVMPHSPAAG	391	
Db	361	GQMGARAGGSLGVLVRPPRPYVMPHSPAAG	391	

RESULT 5

```

RES001.3
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

```

```

Query Match          99.7%; Score 1944; DB 4; Length 729;
Best local similarity 99.7%; Pred. No. 1.7e-153;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Qy 1 MVDVFGALPPPEINARMYVAGPGASLVAQAQWDSVADLFSAAAFQSVVMGLTVGSWIG 60
Db 142 MVDVFGALPPPEINARMYVAGPGASLVAQAQWDSVADLFSAAAFQSVVMGLTVGSWIG 201

```



```
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-819-111-106

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINARMYAGPGSASLVAAAQWMDVSASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDGALPPEINARMYAGPGSASLVAAAQWMDVSASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAASAPYVAMSVTAQAEALTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAASAPYVAMSVTAQAEALTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAAVEEALDTAAANQLMNNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAVMS- - - -LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 IVSMLNNHVSMTNSGVSMAS*THSMLKGFAP-AAAQAVETAQNGVQAMSSLSGLGSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQOMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAAQTPAGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 GLPVGQMGARAG--GGLSGVLVPRPPYVMPHSPAAG 391
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 GLPLQLTNSGGFGGVSNALRMPRAYVMPRPAAG 396
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-8031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids

TREATM
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINARMYAGPGSASLVAAAQWMDVSASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVDGALPPEINARMYAGPGSASLVAAAQWMDVSASDLFSAASAFQSVVWGLTGVSWIG 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SSAGLMVAASAPYVAMSVTAQAEALTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SSAGLMVAASAPYVAMSVTAQAEALTAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LLEQAAVEEALDTAAANQLMNNVPOALQQAQPTGTTSSKLGGLWKTVPSPHSPISN 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAVMS- - - -LGSSL 296
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 IVSMLNNHVSMTNSGVSMAS*THSMLKGFAP-AAAQAVETAQNGVQAMSSLSGLGSSL 299
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQOMLG 356
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAAQTPAGHMLG 359
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 357 GLPVGQMGARAG--GGLSGVLVPRPPYVMPHSPAAG 391
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 GLPLQLTNSGGFGGVSNALRMPRAYVMPRPAAG 396
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
```

```

; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-106

Query Match      84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.3e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY   1 MVDFGALPPEINSARMYAGPGSASLVAAAKWMDVSADLFSASAFAQSVVWGILTVGSWIG 60
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   61 SSAGLMVAASPYYAVNWSVTGAQELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   61 SSAGLMVAASPYYAVNWSVTGAQELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   121 LIATNLGGNTPAIAVNEAEYGEWMWAQDAAMFGYAATAATATATLLPEEAPEMTSAGG 180
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQTCTTPSSKLGGLWKTVSPHRSPLSN 240
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   181 LLEQAVAVEEAIDTAAANQLMNNVPOALQQAQTKSIWPFQDLSLWKAISPHLSFLSN 240
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   241 MYSMANNHNSMTNSGVSTNTLSMWLKGFPAPAAAQAVOTAAQNGVRAMSS---LGSSL 296
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   241 IVSMNNHNSMTNSGVSTNTLSMWLKGFPAPAAAQAVOTAAQNGVRAMSS---LGSSL 299
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   297 GSSGLGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGLG 356
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   300 GSSGLGAGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAQAPGHMLG 359
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   357 GLPVCGMGARAG--GCLSGVLRVPRPYMPHSPPAG 391
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   360 GLPLGLQLTNSGGFGGVSNALEMPPRAYVPRVPAAG 396
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

REFERENCE/DOCKET NUMBER: 210121.411C6
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-112-109

Query Match      76.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.2e-116;
Matches 309; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY   1 MVDFGALPPEINSARMYAGPGSASLVAAAKWMDVSADLFSASAFAQSVVWGILTVGSWIG 60
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   61 SSAGLMVAASPYYAVNWSVTGAQELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   61 SSAGLMVAASPYYAVNWSVTGAQELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   121 LIATNLGGNTPAIAVNEAEYGEWMWAQDAAMFGYAATAATATATLLPEEAPEMTSAGG 180
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   121 LIATNLGGNTPAIEANQAAYSQWNGQDAEAMYGVAATAATATEALLPFEDAPLIITNPGG 180
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQTCTTPSSKLGGLWKTVSPHRSPLSN 240
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   181 LLEQAVAVEEAIDTAAANQLMNNVPOALQQAQVGVPSSKLGGLMTAVSPHLSPLSN 240
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   241 MYSMANNHNSMTNSGVSTNTLSMWLKGFPAPAAAQAVOTAAQNGVRAMSS---LGSSL 296
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   241 VSSIANNHNSMTNSGVSTNTLSMWLKGFPAPAAAQAVOTAAQNGVRAMSS---LGSSL 299
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   297 GSSGLGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMGLG 356
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   300 GSSGLGAGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAQAPGHMLG 359
Db   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
US-08-818-111-104
; Sequence 104, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-104

Query Match 76.3%; Score 1486.5; DB 4; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.2e-116; Indels 5; Gaps 2;
Matches 303; Conservative 16; Mismatches 36;

QY 1 MVDGALPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
D 1 VVDGALPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPVVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
D 61 SSAGLMVAASPVVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTAIAVNEAEYEGEMWAQDAAMFGYAAATATATATLLPPEAPMTSAGG 180
D 121 LIATNLGONTAIAVNEAEYEGEMWAQDAAMFGYAAATATATATLLPPEAPMTSAGG 180
QY 181 LLEQAAVBEASDTAAANGLMNVNPAALQQLAQTGTTTSSKLGKGLWKTVPSPHSPI 240
D 181 LLEQAAVBEASDTAAANGLMNVNPAALQQLAQTGTTTSSKLGKGLWKTVPSPHSPI 240
QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS----LGSSL 296
D 241 VSSIANNHSMNTSGVSMNTLSSMLKGLAP-AAQAVETAENGWAMSSILGSGSSL 299
QY 297 GSGGLGGVAANLGRAASVGSLSVPOWAAAQAQVTPAARALPLTSLTSAABRGQMLG 356
D 300 GSGGLGGVAANLGRAASVGSLSVPPWAAAQAQVTPAARALPLTSLTSAQAQTPCHMLG 359

RESULT 14
US-08-618-112-91
; Sequence 91, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-91

Query Match 60.9%; Score 1187; DB 3; Length 263;
Best Local Similarity 99.6%; Pred. No. 3.9e-91;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 74 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 133
D 1 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
QY 134 IAVNEAEYEGEMWAQDAAMFGYAAATATATATLLPPEAPMTSAGGLLEQAAVBEASD 193
D 61 IAVNEAEYEGEMWAQDAAMFGYAAATATATATLLPPEAPMTSAGGLLEQAAVBEASD 120
QY 194 TAAANGLMNVNPAALQQLAQTGTTTSSKLGKGLWKTVPSPHSPISNMYSMANNHSMNT 253
D 121 TAAANGLMNVNPAALQQLAQTGTTTSSKLGKGLWKTVPSPHSPISNMYSMANNHSMNT 180
QY 254 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 313
D 181 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
QY 314 SV 315
D 241 SV 242

RESULT 15
US-08-818-111-92
; Sequence 92, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

Search completed: November 21, 2003, 16:11:39
Job time : 9.24558 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 15.2499 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672a-26

Perfect score: 1949

Sequence: 1 MVDFGALPPEINSARMYAGP.....SGVLVRPFRPYVMEHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	12	US-10-084-843-107
2	1949	100.0	391	12	US-10-193-002-102
3	1949	100.0	391	12	US-10-088-732A-14
4	1949	100.0	596	9	US-09-287-849-26
5	1949	100.0	596	12	US-10-359-460-26
6	1949	100.0	596	12	US-10-098-732A-20
7	1949	100.0	600	9	US-09-287-849-22
8	1949	100.0	600	12	US-10-359-460-22
9	1949	100.0	729	12	US-10-098-732A-18
10	1949	100.0	930	12	US-10-098-732A-65
11	1944	99.7	729	9	US-09-287-849-2
12	1944	99.7	729	12	US-10-359-460-2
13	1944	99.7	729	12	US-10-098-732A-16
14	1652.5	84.8	396	12	US-10-084-843-111
15	1652.5	84.8	396	12	US-10-193-002-106

16	1486.5	76.3	359	12	US-10-084-843-109
17	1486.5	76.3	359	12	US-10-193-002-104
18	1187.5	60.9	358	9	US-09-287-849-8
19	1187.5	60.9	358	12	US-10-359-460-8
20	1187	60.9	263	12	US-10-084-843-91
21	1187	60.9	263	12	US-10-193-002-92
22	1187	60.9	263	12	US-10-098-732A-12
23	766.5	39.3	400	9	US-09-073-009-126
24	766.5	39.3	400	9	US-09-793-306-126
25	731	37.5	421	16	US-10-080-170-146
26	604	31.0	710	9	US-09-287-849-16
27	604	31.0	710	12	US-10-359-460-16
28	604	31.0	710	12	US-10-098-732A-49
29	604	31.0	856	9	US-09-287-849-12
30	604	31.0	856	12	US-10-359-460-12
31	603	30.9	423	9	US-09-073-009-142
32	603	30.9	423	9	US-09-793-306-142
33	603	30.9	423	12	US-10-098-732A-31
34	588	30.2	394	10	US-09-713-363-205
35	439.5	22.6	655	10	US-09-713-363-207
36	424.5	21.8	943	10	US-09-996-634-131
37	424.5	21.8	943	11	US-09-997-182-131
38	424.5	21.8	943	11	US-09-997-181-131
39	424	21.8	141	9	US-09-073-009-15
40	424	21.8	141	9	US-09-023-588-15
41	424	21.8	141	9	US-09-793-306-15
42	413	21.2	597	9	US-09-793-306-146
43	364.5	18.7	408	16	US-10-080-170-57
44	317	16.3	371	9	US-09-791-171-92
45	317	16.3	371	12	US-09-804-980-92

ALIGNMENTS

RESULT 1

US-10-084-843-107
; Sequence 107, Application US/10084943
; Publication No. US20030143243A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843

FILING DATE: 25-Feb-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.


```
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-098-732A-14

Query Match      100.0%; Score 1949; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 7.1e-146;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLMVAAAAPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300
DB 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGPGOMLGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGPGOMLGLPV 360

QY 361 GQMGARAGGGLSGVLVRPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVRPPRYVMPHSPAAG 391

RESULT 4
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596

; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-098-732A-14

Query Match      100.0%; Score 1949; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 7.1e-146;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLMVAAAAPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300
DB 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGPGOMLGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGPGOMLGLPV 360

QY 361 GQMGARAGGGLSGVLVRPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVRPPRYVMPHSPAAG 391

RESULT 5
US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match      100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 60
DB 9 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 68

QY 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 120
DB 69 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 128

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
DB 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 188

QY 181 LLEQAAAVEASDPTAAANQLMNVPOALQQAQPTQCTTPSSKLGGLWKTVPSPISPIN 240
DB 189 LLEQAAAVEASDPTAAANQLMNVPOALQQAQPTQCTTPSSKLGGLWKTVPSPISPIN 248

QY 241 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVTAAQNGVRAMSSLGSLGSSG 300
DB 249 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVTAAQNGVRAMSSLGSLGSSG 308

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTAAALPLTSLTSAERPGQMLGGLPV 360
DB 309 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTAAALPLTSLTSAERPGQMLGGLPV 368

QY 361 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 391
DB 369 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 399

RESULT 6
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; TYPE: PRT
; LENGTH: 596
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
US-10-098-732A-20

Query Match      100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 60
DB 9 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 68

QY 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 120
DB 69 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 128

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
DB 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 188

QY 181 LLEQAAAVEASDPTAAANQLMNVPOALQQAQPTQCTTPSSKLGGLWKTVPSPISPIN 240
DB 189 LLEQAAAVEASDPTAAANQLMNVPOALQQAQPTQCTTPSSKLGGLWKTVPSPISPIN 248

QY 241 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVTAAQNGVRAMSSLGSLGSSG 300
DB 249 MVSWMNNHMTNSGVSMNTLSSMLKGFAPAPAAAQAVTAAQNGVRAMSSLGSLGSSG 308

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTAAALPLTSLTSAERPGQMLGGLPV 360
DB 309 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTAAALPLTSLTSAERPGQMLGGLPV 368

QY 361 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 391
DB 369 GQMGARAGGSLGVLRVPRPYVMPHSPAAG 399

RESULT 7
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match      100.0%; Score 1949; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.2e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 60
DB 9 MVDFGALPPPEINSGARMYAGFGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSMIG 68

QY 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 120
DB 69 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI 128
```


Db 442 LGGVAA NLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSA AERCGQMLGGLPV 501

Qy 361 GQMGRAGGGLSGVLRVPPRPVYMPHSPAAG 391
|||
Db 502 GOMGARAGGGLSGVLRVPPRPVYMPHSPAAG 532

RESULT 10

```

US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TGA or MAPS)
US-10-098-732A-65

```

Query Match	100.0%;	Score 1949;	DB 12;	Length 930;
Best Local Similarity	100.0%;	Pred. No. 2.2e-145;		
Matches 391;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MYDFGALPPEINSAARMYAGPSASLVAAAQWMDLSVADLFSAAASAFQSVVWGLTVGSWIG	60
Db	142	MYDFGALPPEINSAARMYAGPSASLVAAAQWMDLSVADLFSAAASAFQSVVWGLTVGSWIG	201
QY	61	SSAGILWAAAAPYVAMNSVTAGQELTAAQVRVAAAAAYETAYGLTVPPPVIAENRAELM	120
Db	202	SSAGILWAAAAPYVAMNSVTAGQELTAAQVRVAAAAAYETAYGLTVPPPVIAENRAELM	261
QY	121	LIATNLLGQNTPAIAVNEAEYGEWMAODAAAMPFYAAATATATATLPLEEAPEMTSAGG	180
Db	262	LIATNLLGQNTPAIAVNEAEYGEWMAODAAAMPFYAAATATATATLPLEEAPEMTSAGG	321
QY	181	LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIN	240
Db	322	LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIN	381
QY	241	MYSMANNHSMNTNSGVSMNTLSSMLZGFAPAAAAQAVQTAAQNGVRAMSSLGSSLSGSSG	300
Db	382	MYSMANNHSMNTNSGVSMNTLSSMLZGFAPAAAAQAVQTAAQNGVRAMSSLGSSLSGSSG	441
QY	301	LGCGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAERGPQOMLGGPLPV	360
Db	442	LGCGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAERGPQOMLGGPLPV	501
QY	361	GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG	391
Db	502	GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG	532

RESULT 11

RESULT II
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1

```

; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-460-2

Query Match          99.7%; Score 1944; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.9e-145;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 202 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 240
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHNNHMTNSGVSMTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSG 300
DB 382 MVSMAHNNHMTNSGVSMTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGSLGVLVPPRPYPVMPHSPAAG 391
DB 502 GQMGARAGGSLGVLVPPRPYPVMPHSPAAG 532

RESULT 13
US-10-098-732A-16
; Sequence 16, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a

```

```

; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
; OTHER INFORMATION: protein MTE72F (Ra12-TB9-Ra35 or MTE32-MTE39)
; OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match          99.7%; Score 1944; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.9e-145;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 202 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 240
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTGTTTPSSKLGGLWKTVPSPHSPISN 381
QY 241 MVSMAHNNHMTNSGVSMTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSG 300
DB 382 MVSMAHNNHMTNSGVSMTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGSLGVLVPPRPYPVMPHSPAAG 391
DB 502 GQMGARAGGSLGVLVPPRPYPVMPHSPAAG 532

```

```

RESULT 14
US-10-084-843-111
; Sequence 111, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

```


Search completed: November 21, 2003, 16:38:13
Job time : 16.2499 secs

This Page Blank (uspio)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1949	100.0	391	2	H70608	probable PPE prote	
2	1656.5	85.0	396	2	H70741	probable PPE prote	
3	1583	81.2	393	2	G70568	probable PPE prote	
4	779.5	40.0	393	2	G70929	probable PPE prote	
5	751	38.5	409	2	A70932	probable PPE prote	
6	737.5	37.8	403	2	H70931	probable PPE prote	
7	736	37.8	423	2	H70931	probable PPE prote	
8	731	37.5	421	2	H87056	PPE-family protein	
9	705	36.2	408	2	G70925	probable PPE prote	
10	702	36.0	391	2	H70625	probable PPE prote	
11	702	36.0	413	2	F70560	probable PPE prote	
12	688.5	35.3	463	2	G70931	probable PPE prote	
13	682.5	35.0	468	2	B70932	probable PPE prote	
14	675	34.6	390	2	A70646	probable PPE prote	
15	668.5	34.3	394	2	H70881	probable PPE prote	
16	667.5	34.2	385	2	G70503	probable PPE prote	
17	635	32.6	350	2	H70929	probable PPE prote	
18	633.5	32.5	365	2	H70929	probable PPE prote	
19	615.5	31.6	402	2	A70882	probable PPE prote	
20	603	30.9	423	2	G70582	probable PPE prote	
21	597.5	30.7	391	2	D70922	probable PPE prote	
22	589	30.2	406	2	E70675	probable PPE prote	
23	588	30.2	394	2	A70504	probable PPE prote	
24	575.5	29.5	391	2	F70663	probable PPE prote	
25	495.5	25.4	3300	2	D70575	probable PPE prote	
26	473	24.3	180	2	G70834	probable PPE prote	
27	452.5	23.2	3716	2	E70969	probable PPE prote	
28	448.5	23.0	580	2	D70570	probable PPE prote	
29	444	22.8	346	2	H70874	probable PPE prote	

Db 361 GQMGARAGGGLGVLPRPPRYMHPSPAAG 391

RESULT 2

H70741
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70741
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70741
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-396 <COL>
A:Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA999966.1; PID:e250360;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 85.0%; Score 1656.5; DB 2; Length 396;
Best Local Similarity 85.1%; Pred. No. 6.4e-91;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;
Qy 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Qy 61 SSAGLMVAASPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Qy 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 180
Qy 181 LLEQAAVEEASDTAAANQNMNVPOALQLOAQPTGCTPSSKLGGLMKTVPSPHRPSIN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQLOAQPTGCTPSSKLGGLMKTVPSPHRPSIN 240
Qy 241 MVSMMNNHSMNTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
Db 241 MVSMMNNHSMNTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
Qy 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLG 356
Db 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLG 356
Qy 357 GLPVGOMGARAG--GGLSGVLPRPPRYMHPSPAAG 391
Db 360 GLPLGQLTNSGGFGVSNALRMPRAYVMPRVPAAAG 396

RESULT 3

C70568
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70568
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70568
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-393 <COL>
A:Cross-references: GB:295390; GB:AL123456; NID:G3261766; PIDN:CAB08702.1; PID:e315074; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 81.2%; Score 1583; DB 2; Length 393;
Best Local Similarity 81.5%; Pred. No. 1.4e-86;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;
Qy 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Qy 61 SSAGLMVAASPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Qy 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 180
Db 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 180
Qy 181 LLEQAAVEEASDTAAANQNMNVPOALQLOAQPTGCTPSSKLGGLMKTVPSPHRPSIN 240
Db 181 LLEQAAVEEASDTAAANQNMNVPOALQLOAQPTGCTPSSKLGGLMKTVPSPHRPSIN 240
Qy 241 MVSMMNNHSMNTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 296
Db 241 VSSIANNHSMNTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMSS----LGSSL 299
Qy 297 GSSGLGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLG 356
Db 300 GSSGLGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLG 359
Qy 357 GLPVGOMGARAGGGLGVLPRPPRYMHPSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPRVPAVAIPRTPAAG 393

RESULT 4

G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAA17711.1; PID:e125460;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 40.0%; Score 779.5; DB 2; Length 393;
Best Local Similarity 43.7%; Pred. No. 4.3e-39;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;
Qy 2 VDFGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 61
Db 1 MVDGALPPEINARMYAGPGSAPMWAAAAMGLAAELSSAAGHETVITQSSGWLGP 60
Qy 62 SSAGLMVAASPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 121
Db 61 ASAAAEAVAPYVAMVSTAGQALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70931
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL02021; GB:AL123456; NID:G3250699; PIDN:CAA17722.1; PID:el25461
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 37.8%; Score 736; DB 2; Length 423;
Best Local Similarity 41.8%; Pred. No. 1.7e-36;
Matches 182; Conservative 50; Mismatches 123; Indels 80; Gaps 12;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIGS 61
DB 1 MFDGULPPEINSGRMYTGFPGPMLAATAWDLAVELHATAAGVASELSALT-GANSGP 59
QY 62 SAGLWVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 60 SSTSVASAAAPYVAMSVATAVFAELAGQARLALAAEFAAFAATVPPPPVIAENRAELMVL 119
QY 122 IATNLLGONTPTAIVNEAEYEGEMWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 181
DB 120 IATNIFGONTPTALMTEAYGEMWQAQDAAMFGYAGGATATA-SRTATETPPTTNHQL 178
QY 182 LEQAAVEASPTAAN-----QIMNVPOALQLOAQT-----QGTP-----S 221
DB 179 GAQSSAVAQTAATAAGNLQSAFFQLLSAVPALQGLAPTASQASATPQWVTDLGNLS 238
QY 222 SKLGLWKTVPSPHRSPISNVSMNHNMTNSGVSMTNTLSSMLKGFAPAAAQVOTA 281
DB 239 TFLGG--AVTGPTTP-----GVLPSPGVPYLLGQSVL-----V 271
QY 282 AQNGVRAMSLGS-----SLGSSGLGG--VAANLGRAASVGSLS 319
DB 272 TQNGQGVSAALLKIGSKPITGALAPLAEFALHTPILSGELGGSVSAGIGRAGLVGKLS 331
QY 320 VQQAANAQVTPAARALPLTSLSS--AAERGQGMQLGGLPVQCMQARAGGLSGVLVR 376
DB 332 VPQGMVAAPEPSPAALQATRLAAAPTATDAGALLGGMALSGLAGRAAAGSTG--- 388
QY 377 VPPRPVMPHSPAAAG 391
DB 389 ---HPIGSAAAPAVG 400

RESULT 8
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eigmsier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squires, R.; S
A:Title: Massive gene decay in the leprosy bacillus
A:Reference number: A85909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:G13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 37.5%; Score 731; DB 2; Length 421;
Best Local Similarity 41.1%; Pred. No. 3.4e-36;
Matches 174; Conservative 55; Mismatches 160; Indels 34; Gaps 8;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
DB 1 MFDFAALSPEITNSTRMYLPGSGSPILTAATAAAMVWLAKELTAAQAQGLQSAVEAL-LTTFEG 59
QY 61 SSAGLWVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
DB 60 ESAAALAEVTPPEKWLTONAASAEITATOLTVAANAYETATFTVPPPLVFNRAQACU 119
QY 121 LIATNLLGONTPTAIVNEAEYEGEMWQAQDAAMFGYAAATATATATLLPPEEAPMTSAG 180
DB 120 LIMSNIFFGQNSTAIAEKEAEYTEMWIQDAAMTSYQASVLEAVGATKATAPPLGVNEVG 179
QY 181 L-----LQAAAVEASPTAANQMLNN-----VPOALQ-----LAOP 214
DB 180 LAQEVVEEVVEEVVEVEVEAEQAISQAALDQAVNEGEATVVPQDQVNDVATP 239
QY 215 TQGTTPSSKLG--LWKTVPSPHRSPISNVSMNHNMTNSGVSMTNTLSSMLKGFAPA 272
DB 240 QTAVPSSSSAAAPQLNGGFAQLHLSPLNDTSLMNNHAGMANAGLSLVNCGSAMKSLAP- 298
QY 273 AAQAQVOTRAQNGVRAMSVLSGLSGSLGGVAAANLGRAASVGSLSVQAAANAQVAT 332
DB 299 TTTTAAEAFKAMGSVQSTGRGLGSSGGHVTQAOLGAASTIGSLRVPQTWTTSQPYT 358
QY 333 PAARALPLTSLTSAAREGPGQML-GGLPVQGM---GARAGGLSGVLVRPPRPVMPHSP 388
DB 359 AATRALSPARKAVATSESAPLLGGGLPAPMPVPGSGSGTGVTALRLQPRAFVMPRP 418
QY 389 AAG 391
DB 419 AAG 421

RESULT 9
G70925
Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:Z74024; GB:AL123456; NID:G3250700; PIDN:CAA98377.1; PID:el301025;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 36.2%; Score 705; DB 2; Length 408;
Best Local Similarity 41.8%; Pred. No. 1.1e-34;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIGS 61
DB 1 MFDGULPPEINSGRMYTGFPGPMLAATAWDLAVELHATAAGVASELSALT-GANSGP 60
QY 62 SAGLWVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 61 AAASMAVAATVYVAVLWLSATAGQAEQAGMQARAAAAAYELAFAMTVPPPPVIAENRAELMIL 120
QY 122 IATNLLGONTPTAIVNEAEYEGEMWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGL 181
DB 121 VATNFFGONTPTAATAEAQYAEWQAQDAAMFYAGGATATAI-ELTPFTTAAVPTTSPAL 179
QY 182 LEQAAA-----VEEASDTAAANQMLNN-----VP---QALQO---LAQTOGTTTSSKL--- 224

```

Db 180 ACQAAATVSSTVPPLATTAAPVQLLQQLSSTSLIPWYALQQLWLAENLLGLTPDNRTIV 239
QY 225 -----GGLWKTVPSPHRSPISNMWSVANNHSMNTSGVSMNTLSSMLKGFAPAAAA 275
Db 240 RLIGISYFDEGL-----LQFEASLAQAATPGTGGAG--DSGSSVLDSWGPPIFA 287
QY 276 QAVQTAAQNGVRAMSSL--GSSLSG-----SGLGGGVAANLGRAASVGSLS 319
Db 288 -----GPRASPSVAGGAVGVQTPQPYWYWALDRSIGSVSAALGKGSAGSLS 338
QY 320 VPOWAAANQAVTPAARALP---LTSLSAAERPGQMLGLPVQMGARAGGSLGVLR 376
Db 339 VPPDWAARWANPAWLPEDDVTALRGTAENA---LLRFPWASAGQSTGGGF--VHK 393
QY 377 VPPRPYVMPHSPAAG 391
Db 394 YGFRFLAVMQRDPFFAG 408

RESULT 10
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:Z95436; GB:AL123456; NID:3261714; PIDN:CAB06873.1; PID:e304546;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 391;
Best Local Similarity 42.3%; Pred. No. 1.66-34;
Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;
QY 2 VDFGALPPEINSMYAGPGSASLVAAQMWDSVASDLFSNASAFQSVVWGLTVGWSIGS 61
Db 1 MDFGALPPEINSMYAGPGSASLVAAQMWDSVASDLFSNASAFQSVVWGLTVGWSIGS 60
QY 62 SAGLMVAAASPYVAMSVTAGCAELTAAQVRVAAAYETAYGLTVPPPVIAENRAELMIL 121
Db 61 ASMAWVAAQPYLAWLITVTAEEAAHAGSQAMASAAAYEAAYMTVPEVVAANKALLAAL 120
QY 122 IATNLLGONTTAAIAVNEAEYEMWAOAAAMFGVAATATATATLTPPEAPENTSAGGL 181
Db 121 VATNVLGINTTAIMATEALYAEWMAQDALAMGYVAAASG--AAGMLQLPLSPSQTNPGL 179
QY 182 LEQAAVEEASDTAAAHQ-----LMNVVPOALQOLAQPTQCTTPSSKLGGLWKTVPSPHRS 236
Db 180 AAQSPAVGSAATAAVNQVSVADLISLPAVSGLASPVTSVLDSTGLSGLIIDALLA 239
QY 237 P--ISNMVSMANNHMS-MTNSGVSMNTLSSMLKGFAPAAAAQAVOTAQNGVRAMSSLG 293
Db 240 TPFVANIINSVNTAAWYVNAIPTAIFLALNLSGAPVIAEAGIAEABEG--AASAAA 296
QY 294 SSLSGSLGGVAAANLGRAASVGSLSVPOWAAANQAVTPAARALPTLSLTSAAERGPQG 353
Db 297 AGLADSVTPAGLGSGLGATLVGLSLVFAAWSTAAATACATALEGSGWTVAAEER--GP 355
QY 354 MLGGLPVQMGARAGGSLGVLRVPPRPYPVP 385
Db 356 VTGNWP--GMASAAKGTGAYAGFRYGFKPTWMP 386

```

RESULT 11

F70560
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: F70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:Z95436; GB:AL123456; NID:3261770; PIDN:CAB08826.1; PID:e315656; I
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match

36.0%; Score 702; DB 2; Length 413;
Best Local Similarity 39.1%; Pred. No. 1.78-34;
Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;
QY 1 MVDFGALPPEINSMYAGPGSASLVAAQMWDSVASDLFSNASAFQSVVWGLTVGWSIGS 60
Db 1 MLDFAQLPPEVNSALMYAGPGSGPMLAAANAEALAELOTTASIVDALITGLADGFWQ 60
QY 61 SSAGLMVAAASPYVAMSVTAGCAELTAAQVRVAAAYETAYGLTVPPPVIAENRAELM 120
Db 61 SSAASMVAAATPQVWLRLSTRAGCAEQAGSQAASAAAYEAFATVPPEIAANRALLMA 120
QY 121 LIATNLLGONTTAAIAVNEAEYEMWAOAAAMFGVAATATATATLTPPEAPENTSAGS 180
Db 121 LIATNLLGONTTAAIAVNEAEYEMWAOAAAMFGVAATATATATLTPPEAPENTSAGS 179
QY 181 LLEQAAVEEASDTAAANQLMNVVPOALQOLAQPTQCTTPSSKLGGLWKTVPSPHRSPI 240
Db 180 LASQAASVQAVSGAANAQALTDIPKAL-----FGLSGIFTNEPPLDGLK 226
QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAA---AAQAVOT----- 280
Db 227 ALGLTGTHTSWSDGSLIVGGVLGDFVQVGTGSAELDASVAMDTFGKWSFARLMTQFKD 286
QY 281 -----AAQNGVRAMSSILSGSLGSGGVAANLGRAASVGSLSVPOA 323
Db 287 YFGLAHDLPKWASEGAKAAGEAKALPAAVPAIPSNGL--SGVAGAVGQAASVGGGLKVP 345
QY 324 WAAANQAVTPAARALPTLSLTSAAERGPQMLGGLPVQMGARAGGSLGVLR--RVPPRP 381
Db 346 WTATTPAASPVALAASNLGALAAAEAGSTHAFGGMPL--MGSGAGRAFNFAAPRYGFP 403
QY 382 VYNPHSPAAG 391
Db 404 TVIAQPPAGC 413

RESULT 12

C70931
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-463 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAAL17730.1; PID:e125462
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;
Best Local Similarity 42.8%; Pred. No. 1.2e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPPPEINARMYAGSGASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 61
DB 1 MDFGVLPPEINSGRMVAGFGSGPMLAAAAAWDGLATELQSTAAADYGSVISVLT-GVNSGQ 59
QY 62 SAGLWVAASPYVAVMSVTTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 60 SSGTVAARAAPVAVMSATAALAREAAQAASAAAAYEAFATVPPVVAANRAELAVL 119
QY 122 IATNLGQNTPAIVNAEAYGEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 181
DB 120 AATNIFGQNTGAIAAEARYAEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 178
QY 182 LEQAAVEEASDTAAANQLMNVVPOALQLOAQPTQGTTPSSKLGGLWKTVS--PHRSPi- 238
DB 179 ATQGVAAQVAGASGN-ARSLVSEVLEFLA--TAGTWNKTVASLMNAVTVGPVASSVY 235
QY 239 -----SNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRA 288
DB 236 NSMLGLGFABSKNVLPAANDTVISTFGVQFKFNPVTPFEDLIPK----- 283
QY 289 MSLSGLSGLG-----SSGLGG-----GVANLGRAASVGSLSVPOAAANQAVTPAARALPL 340
DB 284 -SALGAGLGRSAISGLGSTAPASAGASQAGSVGMSVPPSWAAATPAITVAAPVSS 342
QY 341 TSITS--AARPGQML-----GGLPVGQMGARAGGSLGVLVR 377
DB 343 TGLQAVPANAISEGSLLSQVALASVAGGALGGAARATGGFLGGGRV 389

RESULT 13
B70932
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70932
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <COL>
A:Cross-references: GB:283867; GB:AL123456; NID:G3261695; PIDN:CABO6278.1; PID:e291015; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 34.6%; Score 675; DB 2; Length 380;
Best Local Similarity 43.1%; Pred. No. 6.1e-33;
Matches 172; Conservative 56; Mismatches 143; Indels 28; Gaps 12;

QY 2 VDFGALPPPEINARMYAGSGASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 61
DB 1 MDFGALPPPEINARMYAGSGASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 60
QY 62 SAGLWVAASPYVAVMSVTTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 61 AASMTVAAPYICWLYTTAEKTCQTAIQAARALAFQAYATLPPVVAANRIQULLAL 120
QY 122 IATNLGQNTPAIVNAEAYGEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 181
DB 121 IATNFFQNTAAIAATEAQVAEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 179
QY 182 LEQAAVEEASDTAAANQLMNVVPOALQLOAQPTQGTTPSSKLGGLWKTVSPHRSPi 238
DB 180 TAQAAVVSQATDPLSL--LIETVQALQALTIPTSPFEDTFDLAIPAGYATVGVTVQDVE 237
QY 239 SNMVM--ANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAASISGLS 296
DB 238 SFVAGTIGAESNLGGLNVGDNPAEVTGPGDFGIGELVSA---TSPGGGVS-AAG-- 289
QY 297 GSSGLGGVVAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSL--TSAERG-PGQ 353

QY 62 SAGLWVAASPYVAVMSVTTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 61 SSAGMAAAAPYVTVWSGTSQAQAADQARAANVAYETAFRAVPPPPVIAENRSOLISL 120
QY 122 IATNLGQNTPAIVNAEAYGEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 181
DB 121 VATNIFGQNTAAIAATEAQVAEMWQADTWMFGYASSATA-SRLTPTTAPPTQTNPSGL 179
QY 182 LEQAAVEEASDTAAANQLMNVVPOALQLOAQPTQGTTPSSKLGGL- 227
DB 180 AGQAAATGQATALASGNTAVTTALSSAAAFQFPDIITLLQGLA--TLSTQVTVLMGQMI 237
QY 228 -----WKTSPHRSPI-SNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQ 276
DB 238 NAIFGTGATTYQNVFVTAANVTIKFTWANDAMSAPNLGMEFKVF-----WQPPAPE 291
QY 277 AVQTAQNGVRAASISGLSGLG-----SSGLGGVVAANLGRAASVGSLSVPOAAANQAV 331
DB 292 IPK-----SSLGAGLGRSGLSAGLAHAASAGLQCANLVGDLSPVPPSWASATPAV 341
QY 332 TPAARALPLTSLTA-AERPGQMLGGLPVGQMGARAGGGL 371
DB 342 RLIVANTLPATSLAAAPATQIPANLLQOMALGSM---TGGAL 379

RESULT 14
A70646
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70646
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <COL>
A:Cross-references: GB:283867; GB:AL123456; NID:G3261695; PIDN:CABO6278.1; PID:e291015; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 34.6%; Score 675; DB 2; Length 380;
Best Local Similarity 43.1%; Pred. No. 6.1e-33;
Matches 172; Conservative 56; Mismatches 143; Indels 28; Gaps 12;

QY 2 VDFGALPPPEINARMYAGSGASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 61
DB 1 MDFGALPPPEINARMYAGSGASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 60
QY 62 SAGLWVAASPYVAVMSVTTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 61 AASMTVAAPYICWLYTTAEKTCQTAIQAARALAFQAYATLPPVVAANRIQULLAL 120
QY 122 IATNLGQNTPAIVNAEAYGEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 181
DB 121 IATNFFQNTAAIAATEAQVAEMWQADAAAFYAAATATATATATLLPPEEAPMTSAGGL 179
QY 182 LEQAAVEEASDTAAANQLMNVVPOALQLOAQPTQGTTPSSKLGGLWKTVSPHRSPi 238
DB 180 TAQAAVVSQATDPLSL--LIETVQALQALTIPTSPFEDTFDLAIPAGYATVGVTVQDVE 237
QY 239 SNMVM--ANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAASISGLS 296
DB 238 SFVAGTIGAESNLGGLNVGDNPAEVTGPGDFGIGELVSA---TSPGGGVS-AAG-- 289
QY 297 GSSGLGGVVAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSL--TSAERG-PGQ 353

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 4.61043 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-26

Perfect score: 1949
Sequence: 1 MVDGALPPRINSARMYAGP.....SGVLRVPPRYVMPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 Y592_MYCTU	Q10813 mycobacteri
3	688.5	35.3	463	1 Y102_MYCTU	O53951 mycobacteri
4	444	22.8	487	1 Y442_MYCTU	P42611 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 Y248_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y096_MYCTU	Q10892 mycobacteri
9	324	16.6	434	1 YU18_MYCTU	P31500 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 YV29_MYCTU	O06246 mycobacteri
12	217.5	11.2	176	1 YV25_MYCTU	O50703 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 Y126_MYCTU	Q50702 mycobacteri
15	150	7.7	864	1 ELS_RAT	Q99372 rattus norv
16	143.5	7.4	730	1 ELS_HUMAN	P15502 homo sapien
17	139.5	7.2	881	1 PRY3_YEAST	P47033 saccharomyc
18	135	6.9	907	1 A180_HUMAN	O50641 homo sapien
19	133.5	6.8	825	1 ICP0_HSV2H	P28284 herpesimp
20	132	6.8	1120	1 STFR_ECOLI	P60702 escherichia
21	132	6.8	2090	1 N214_HUMAN	P35658 homo sapien
22	131	6.7	836	1 Y226_BPML5	Q05233 mycobacteri
23	131	6.7	1150	1 AFMU_PIG	P12021 sus scrofa
24	129	6.6	1783	1 RAA3_CHLRE	Q9fec4 chlamydomon
25	128	6.6	779	1 SRP_DROME	P52172 drosophila
26	127.5	6.5	790	1 ANP_NOTCO	P24856 notochenia
27	126	6.5	354	1 YAU6_SCHPO	Q10169 schizosacch
28	125.5	6.4	1211	1 BUN2_DROME	Q44523 drosophila
29	125	6.4	577	1 CST2_HUMAN	P33240 homo sapien
30	124	6.4	2090	1 HFC1_MESAU	P51611 mesocricetu
31	123.5	6.3	677	1 Y136_MYCTU	Q50597 mycobacteri
32	123	6.3	432	1 YF10_MYCTU	P17189 mycobacteri
33	123	6.3	1140	1 YN96_YEAST	Q04893 saccharomyc

34	122	6.3	394	1 HYF1_ALCEU	P45805 alcaligenes
35	121.5	6.2	1025	1 SLAP_CAUCR	P35828 caulobacter
36	121.5	6.2	1845	1 Z236_HUMAN	Q9u136 homo sapien
37	121	6.2	1199	1 P121_RAT	P52391 ratus norv
38	119.5	6.1	635	1 HMLA_DROME	P10105 drosophila
39	119	6.1	915	1 A180_RAT	Q05140 rattus norv
40	118.5	6.1	580	1 EXPR_XANCP	P23314 xanthomonas
41	118.5	6.1	2038	1 FSH_DROME	P13709 drosophila
42	118	6.1	1508	1 BCSC_XANAC	P58938 xanthomonas
43	117	6.0	444	1 Y808_CHLPN	Q92798 chlamydia p
44	117	6.0	652	1 PICA_HUMAN	Q13492 homo sapien
45	117	6.0	774	1 STP_LAMBD	P03764 bacterioph

ALIGNMENTS

RESULT 1					
YD61_MYCTU					
ID	YD61_MYCTU	STANDARD;	PRT;	396 AA.	
AC	Q11031;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical PPE-family protein Rv1361c.				
GN	Rv1361C OR MT1406 OR MTC102B10.25C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Actinobacteria; Actinomycetales; Actinomycetales;				
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean M.A., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Bishai W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; Z15555; CAA99966.1; -				
DR	EMBL; A5007013; AAK45669.1; -				
DR	PIR; H70741; H70741.				
DR	TIGR; MT1406; -				
DR	TubercuList; Rv1361c; -				
DR	InterPro; IPR000030; Microbac_PPE.				
DR	Pfam; PF00823; PPE; 1.				


```

OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
EX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AL022021; CAAL7723.1; -
CC EMBL; AE007044; AAK46123.1; -
CC PIR; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; Rv1802; -
CC InterPro; IPR00030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 401 401 S -> L (IN REF. 2).
SQ SEQUENCE 463 AA; 46021 MW; EB64828BF09FA551 CRC64;
Query Match
Best Local Similarity 35.3%; Score 688.5; DB 1; Length 463;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;
QY 2 VDFGALPPEINSARMYAGPGSASLVAAQAQWDSVADLFSASAFQSVWGLTVGWSIGS 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MDFGVLPPEINSGRMVYAGPGSGPLMAAAAADWGLATELQSTADYDGSVSLT-GVWSGQ 59
QY 62 SAGLVAAASPYVAWMSVTAGQELTAQVRAAAVETAGLVTPPPVTAENRAELMIL 121
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 SSGVMAAAAPYVAWMSATALEAAEAQAASAAAYEAFAATVPVVAANRAEAVL 119
QY 122 IATNLGQNTPAIAVNEAYGEMWAQDAAMFGYAAATATATATLLPFEAPETSAGGL 181
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 AATNIFGQNTCAIAAEAAEARYEMWAQDAAMFGYAGSSVAT-QVTPFAAPPTTNAAGL 178
QY 182 LEQAAVEAEASDTAAANOLMNVQALCOLAQPTCGCTTPSKLGLKWTYS--PHRSP1- 238
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 ATQGVAAQAQVAGSAGN-ARSLSEVLEFLA--TAGTNYNKTVASLNMVNTGVPIYASVY 235
QY 239 -----SNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRA 288
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 NSMLGLGPAESKWLIPANDTVISIFGVQKQFENFVTPNPDLIPK----- 283
QY 289 MSSLGSSLG-----SSGLGG---GVANLGRAASVGLSVFQVAAVAAANQAVTPAARALPL 340

```

```

Db 284 -SALGAGLGRSAISSGLGSTAPISAGASQAGSVGMSVPSMAAATTAIRTVAAVFS 342
QY 341 TSLTS--AARPGQML-----GGLPVGMGARAGGLSGVLRV 377
Db 343 TGLQAVPAAAISEGLLSQMALASVAGGALGGAARATGCGFLGGGRV 389
RESULT 4
Y442_MYCTU
ID Y442_MYCTU STANDARD; PRT; 487 AA.
AC P42611; O53727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein Rv0442c.
GN Rv0442c OR MT0458 OR MT037.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
RL J. Bacteriol. 169:1080-1088(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; M15467; AAA88235.1; ALT_INIT.
CC EMBL; AL021932; CAAL7399.1; -
CC EMBL; AE006948; AAK44681.1; -
CC PIR; C70830; C70830.
CC TIGR; MT0458; -
CC TubercuList; Rv0442c; -
CC InterPro; IPR00030; Microbac_PPE.
CC InterPro; IPR002989; Mycobac_pentapep.
CC Pfam; PF01469; Pentapeptide_2; 5.

```



```
RESULT 6
VF48 MYCTU
ID_YF48 MYCTU STANDARD; PRT; 678 AA.
AC Q10776;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein Rv1548c.
GN Rv1548C OR M1599 OR M1599.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gencles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z74020; CAA98335.1; -;
DR EMBL; AE007026; AAK45866.1; ALT_INIT.
DR FIR; A70762; A70762.
DR TIGR; MT1599; -.
DR Tuberculist; Rv1548c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 11.
DR Pfam; PF08823; PPE; 1.
DR Pfam; PF08823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34
FT TRANSFAM 180 200
FT TRANSFAM 258 258
FT CONFLICT 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
Query Match 21.4%; Score 418; DB 1; Length 678;
Best Local Similarity 31.9%; Pred. No. 2.9e-17;
Matches 106; Conservative 48; Mismatches 140; Indels 38; Gaps 7;
QY 2 VDFGALPEINSARMYAGGSGASLVAAQWDSVASDLFSAASQSVVWGLTVGWSIGS 61
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MNFSLVLPPEINSALMFAGAGGCPMLAASANTGLAGDLGSAASFSATVTSQATGWSQGP 60
QY 62 SAGLMVAASPYVAMSVTAGQAEILTAQVRVAAAAYETAVGLTVPPPPVIAENRAELMIL 121
```

```

CC -----
DR EMBL; U00015; AAC43220.1; -
DR EMBL; X68431; CAA48480.1; -
DR EMBL; Z21952; CAA99950.1; -
DR EMBL; Z97179; CAB09938.1; -
DR EMBL; AL583918; CAC29919.1; -
DR PIR; C86960; C86960.
DR PIR; S33522; S33522.
DR PIR; S39872; S39872.
DR Leproma; ML0411; -
DR InterPro; IPR000030; Microbac_PPE.
DR Antigen; Repeat; Complete proteome.
KW Antigen; Repeat; Complete proteome.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
FT REPEAT 209 214 1.
FT REPEAT 230 235 2.
FT CONFLICT 132 132 T -> S (IN REF. 2).
FT CONFLICT 189 189 S -> L (IN REF. 2).
FT CONFLICT 191 191 H -> D (IN REF. 2).
FT CONFLICT 292 292 P -> L (IN REF. 2).
SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BEC0D6E6A9D8 CRC64;

Query Match
Best Local Similarity 18.7%; Score 364.5; DB 1; Length 408;
Matches 115; Conservative 72; Mismatches 163; Indels 97; Gaps 11;

QY 1 MVDFGALPPEINARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSGWG 60
Db 1 MEDFMVSPVNAFLMRGPGSTPLWGAAEAMISLAQLMEAAQVSDTVIVAVPASPAG 60
QY 61 SSAGLMVAASPVMVMTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 ETSMDLASRSTVFAVLDCGNAENAGLIARVLHAYAFEPERAGMVPVLLTVLGNITHTMA 120
QY 121 LIATNLGQNTPAIAVNEAIEYEMWAQDAAMFGVAAATATATATLPPFEAPEMTSAGG 180
Db 121 LXAINWFGQVSTVALEADYDLWLVQNSTMTTTRDTVLRGKXENFEPAPQLVSR-Y 179
QY 181 LLEQAAVBEASDTAAANGMNNVPAQLQLAQ-
Db 181 CMDRDSVNSFHSSSSSDSLYSIDNLYDVAQSEHSGSDMSQSYNTCGSVAQSELCD 239
QY 214 ----PTQ-----GTFSSKLGGLKWTVPSPINNVSMANNHMTNMTNMTNMTNMTNMTN 261
Db 240 PFGTPSQSQNDLSATSLTQGLG-----DSIISSASLTTN-----ISSST 286
QY 262 LGSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSGGGVAAAN-----LGRAASVG 316
Db 287 ASSIM----PIVASQVTEITLGRSQV-AVEKMTIQISSTAVSDVAASKVWAGVQAVSVG 341
QY 317 SLSPVQAAANAQVTPARALP--LTSLSAAERPGQMLGGLPVQGMGARAGGLSGV 374
Db 342 ALRVPENWATASQPVNATASHVAPAGCSAITTA-----VSGFLEGV 381
QY 375 LRVPFRPYVMPHSPAG 391
Db 382 TQ--PAEEVLITASVAGG 396

RESULT 8
Y096 MYCTU STANDARD; PRT; 463 AA.
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV0096.
GN RV0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;

```

```

[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74410; CAA89932.1; -
DR EMBL; AE006922; AAK4327.1; -
DR PIR; H70750; H70750.
DR TIGR; MT0105; -
DR TuberculList; RV0096; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE, 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 46894 MW; 42D9D6GA033D0DD8 CRC64;

Query Match
Best Local Similarity 16.6%; Score 324.5; DB 1; Length 463;
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;

QY 6 ALPPEINARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSGWG 65
Db 2 AIPPEVHSGLLSAGCGPGSLLVAAQWQELSDQYALACELGQLGGEVQASSWQGTAAQ 61
QY 66 MVAASPPYVAVMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMIATN 125
Db 62 YVAAGFYLANLEQTAINSAVTAAGVAAAAYCSLAAMPFAELAAHAIHGVLIATN 121
QY 126 LQGNTPTAIVNEAIEYEMWAQDAAMFGVAAATATATATLPPFEAPEMTSAGGLLEQA 185
Db 122 FFGINTVPVIALNEADYVRVWMLQAADTMAAYQAVADAATVAVPSTQPPAPPGG---- 176
QY 186 AAVEEASDT-----AAANQIMNNVPAQLQLAOPTQ-----GTTFSSKL----- 224

```



```
Db 177 ----DAADTRLDVLSIGQLIRDI---LDFIANPKYFLEFFEQFGSPAVTVVLAVAL 229
Qy 225 ----GGIM-----KTVSPHRSPISNMVMNHNMTSGVSNMTLSMLK-----GF 269
Db 230 QLYDPLWPPYASYGLLLPFTTP-----TSLALTALSALIHLLNLPAGL 275
Qy 270 APAAQAQVQTAQNGVRAMSSGLSSGLGGVAAANLGRAASVGLSVLPQAWAANQ 329
Db 276 LPFAAAA-----LPGDQWGANLAVATPATAAVP-----GGSP 308
Qy 330 AVTPAARALPLTSLTSAERPG-----QMLGGLPVG-QMGARAG 368
Db 309 PTSNPAPAPASNSVGSASAPGISYAVPGLAPGVSSGPRAG 351

RESULT 9
YU21_MYCTU
ID YU21_MYCTU STANDARD; PRT; 434 AA.
AC P31500; O53265;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical PFE-family protein RV3021c.
GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Taylor K., Whitehead S.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 160-374 FROM N.A.
RC STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AL021287; CAA16103.1; -
DR EMBL; AE007129; AAK47427.1; ALT SEQ.
DR EMBL; AE007129; AAK47430.1; ALT SEQ.
DR EMBL; X59271; CAA41961.1; ALT_FRAME.
DR PIR; E70857; E70857.
DR TIGR; MT3098; -
DR TIGR; MT3101; -
DR TubercuList; RV3018c; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;

Query Match 16.8%; Score 324; DB 1; Length 434;
Best Local Similarity 28.0%; Pred. No. 4.7e-12;
Matches 115; Conservative 56; Mismatches 178; Indels 62; Gaps 12;

Qy 6 ALPPEINSARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGSAGL 65
Db 8 ASPPEVHSALLSAGPGPGSLQAAAAGWSALSIEYAAVAQELSVVVAAGVWQGSAGL 67
Qy 66 MVAASPYVAMSVTAGQAELTAAQVRVAAAYETAYGLUTVPPVIAENRAELMILIATN 125
Db 68 FVAAYVYVAVLQAGADGAAAGAAAGVVCALAEWPTLPFLAANHLTHAVLATN 127
Qy 126 LLQONTPTATVNEAEYCEMKAQDAAAFVGAATATATATLTPFEAPEMTSAGLLEQA 185
Db 128 FFGINTIPALNEADYVRMVQATVNSAYEAVVGAALVATPHGTGPAPVIVKEG 181
Qy 186 AAVEEASDTAAAN-----QLMNVVPAQLQLAQTOGTTPTSSKLGGLWKTVSPH 234
Db 182 --ANEASNAVAAATITFPFHEIVQLEETFAAYDQVLSALLSELP--VAWVWFOLFVD 237
Qy 235 ---RSFISNMVSNMHNMTNSGVMTNTLSMLKGP----- 270
Db 238 ILGNFIIGFIITLASNAQLTFEAINASVAVGLYAIAGVIDIVVEWIGNLFGVVP 297
Qy 271 -----PAAQAQVQTAQNGVRAMSSGLSSGLGGVAAANLGRAASV-GSLSVPOA 323
Db 298 GGPLLGAALAAVVPVGVAGLVAGLAAL-PAVGA--AGAPALVGSVAPVSGVSVPOA 354
Qy 324 WAAANQAVTPAARALPLTSLTSAERPGQMLGGLPVGQNGARAGGLSGV 374
Db 355 RLVS--AVEPAPASTSVSVL--ASDRGAGAL--GF-VGTAGRESVGQAGL 398

RESULT 10
YU21_MYCTU
ID YU21_MYCTU STANDARD; PRT; 435 AA.
AC O53268; O53269;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV3021c/RV3022c.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Taylor K., Whitehead S.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 160-374 FROM N.A.
RC STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 82.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL021287; CRA16106.1; ALT_FRAME.
 DR EMBL; AL021287; CRA16107.1; ALT_FRAME.
 DR EMBL; AB007129; AAK47435.1; -.
 DR TIGR; MT3106; -.
 DR TubercuList; Rv3021c; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 299 299 G -> A (IN REF. 2).
 FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
 FT CONFLICT 326 326 L -> V (IN REF. 2).
 SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;

Query Match 16.5%; Score 321.5; DB 1; Length 435;
 Best Local Similarity 26.5%; Pred. No. 6.6e-12;
 Matches 115; Conservative 53; Mismatches 159; Indels 107; Gaps 12;
 QY 6 ALPPEINARMVAGPGSASLVAAQWDSVADLFAASAFQSVVWGLTVGWSAGL 65
 DB 8 ASPPEVHALLSAGPGSLQAAAGWSALSAYAAQELSVMVAAGVWQGPSREL 67
 QY 66 MVAASPYVWMSVTAGQELTAAQVRAAAAYETAYGLTVPPPPVIAENRAELMILITN 125
 DB 68 FVAAYVPYVWLVQASADSAAGAAAGEHEAAAGVVCALAEPTLPBLAAHNLTHAVLVATN 127
 QY 126 LLGQNTPAIVNEAEYGEWQDAAAMFCYAA----- 157
 DB 128 FFGINTIPALNEADTVRMVQQAIVMSYAEVVGAAVLVATHTGPAPVIVKPGANEASN 187
 QY 158 ATATATATALLPPEE-----APEMTSAGLLEQAAYVEEASDTAAANQLMNNVPQALQ 209
 DB 188 AVAAATITPPFGELAKFLEMAAQAPTEVGLIMKSAEAWGVFELITGLVNFEP---- 243
 QY 210 QLAQPTQGTTPSKLGLLWKTVPSPHSPISNMVSM-----ANNH 248
 DB 244 -----LVIV-----LTGMIDMFFATVGFALGVFLVLPFLLEFAVVLE 278
 QY 249 MSMTNSGVSWNTLTSS-----MLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSGL 301
 DB 279 LAITSIGLIISNIFGAIPVLGPGLLGALAAAVVPGVAGLVAGLAALPAVGAAGAP-- 336
 QY 302 GGGVAANLGRASV--CSLSVPQAAWAAANQAVTPAARALPLTSLTAAERGPGQMGLGLPV 360
 DB 337 -----AALVGSVAPVSGGVSPQARLVLS--AVEPAPASTSVSYL--ASDRGAGAL--GF-V 385
 QY 361 GQMGARAGGGLSGV 374

DB 386 GTAGKSVGQGPAGL 399
 RESULT 11
 YI29 MYCTU
 ID YY29 MYCTU STANDARD; PRT; 178 AA.
 AC O06246;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical PPE-family protein Rv3429.
 GN Rv3429 OR MT3533 OR MYCY77.01.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z95389; CAB08678.1; -.
 DR EMBL; AE007158; AAK47873.1; -.
 DR TIGR; C70975; C70975.
 DR TubercuList; Rv3429; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;
 Query Match 12.0%; Score 233.5; DB 1; Length 178;
 Best Local Similarity 35.9%; Pred. No. 3e-07;
 Matches 65; Conservative 25; Mismatches 76; Indels 15; Gaps 3;
 QY 7 LPPEINARMVAGPGSASLVAAQ-----WYDS---VASDLFSAASAFQSVWGLTVGSI 59
 DB 5 IPPEINIIIEGPGADUSAAAAQRLMYNSANMTAKSLTDLGLQBE-----NWK 56
 QY 60 GSSAGLMMVAASPYVWMSVTAGQELTAAQVRAAAAYETAYGLTVPPPPVIAENRAELM 119

Db 57 GSSDLMAAGRYLDMTKHSRQILETAYVIDELAVYVEETRHKVVPATIANNREEVH 116
 QY 120 ILIATNLGONTPALAVNEAEYGEWMAODRAAMFGYAAATATATATLLPPEEAPENTSAG 179
 Db 117 RLIASNAGVNTPALAGLDAQYQYTRAGNIAMVNDYQSTARFILAYLPRWQEPPIYGGG 176
 QY 180 G 180
 Db 177 G 177

RESULT 12

Y125 MYCTU
 ID Y125 MYCTU STANDARD; PRT; 176 AA.
 AC Q50703;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical PPE-family protein RV3425.
 GN RV3425 OR MTCY78.04C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z77165; CAB01031.1; -
 CC PIR; F70738; F70738.
 CC TubercuList; RV3425; -
 CC InterPro; IPR000030; Microbac_PPE.
 CC Pfam; PF00823; PPE; 1.
 CC KW Hypothetical protein; Complete proteome.
 CC SEQUENCE 176 AA; 19855 MW; B8CFE2E9463B87B0 CRC64;

Query Match 11.2%; Score 217.5; DB 1; Length 176;
 Best Local Similarity 33.5%; Pred. No. 2.5e-06;
 Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;
 QY 7 LPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGLM 66
 Db 5 IPAEYISNIIEYFGGADSLFFASGQLRELAYSVETTAESLEDELD-ENWKGSSDLL 63
 QY 67 VAASPYVAMSVTAGQBELTAAQVRAAAAYETAYGLTVPPPVIAENRAELMILITNL 126
 Db 64 ADAVERYLQWLKSKHSQKHAHWINGLANYNDRKRVWPPETIANRERRELIASNV 123
 QY 127 LGONTPALAVNEAEYGEWMAODRAAMFGYAAATATATATLLPPEEAPENTSAG 179
 Db 124 AGVNTPALADLDAQDQVRRARVAVNAYVSWTRALSLSLDPREWPPEPPIYRG 176

RESULT 13

ELS_MOUSE
 ID ELS_MOUSE STANDARD; PRT; 860 AA.
 AC P54320;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastin precursor (Tropoelastin).
 GN ELN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Lung;
 RX MEDLINE=95130069; PubMed=7829060;
 RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
 RT "Use of an intron polymorphism to localize the tropoelastin gene to
 RT mouse chromosome 5 in a region of linkage conservation with human
 RT chromosome 7."
 RL Genomics 23:125-131(1994).
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U08210; AAA80155.1; -
 CC PIR; A55721; EAMS.
 CC MGD; MGI:95317; Eln.
 CC InterPro; IPR003979; tropoelastin.
 CC PRINTS; PR01500; TROPQELASTIN.
 CC KW Structural protein; Repeat; Signal; Connective tissue.
 CC SIGNAL 1 27 POTENTIAL.
 CC FT CHAIN 1 860 ELASTIN.
 CC FT DISULFID 850 855 BY SIMILARITY.
 CC SEQUENCE 860 AA; 71955 MW; 0C0B55AAE1EDD7F1 CRC64;

Query Match 7.9%; Score 153.5; DB 1; Length 860;
 Best Local Similarity 24.4%; Pred. No. 0.069;
 Matches 115; Conservative 34; Mismatches 192; Indels 131; Gaps 17;
 QY 5 GALP-----PEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSW-- 58
 Db 402 GGIPGVGGPGICGGPGIVGGPGAVSPAATAAKAAKAAK--YGAEGGVGITYGVGAGGFP 459
 QY 59 --IGSAGLNVAAASPYVAMSVTAGQBELTAAQVRAAAAYETAYGLTVPPPVIAENRA 116
 Db 460 YGVGAGAGL--GGASPA--AAAAAAAAAAAAAAAAAAAAAAAAAAKAAKYG-----AGGAG 491
 QY 117 ELMILITNLGQNTFAIVNEAEYGEWMAQDAAMFGVAAATATATATATLLP-FEEAPEM 175
 Db 492 ALGGLVPGAVPGALPGAVPAVPGAGVPGAGTFAAAAAA--KAAKAAKAGLPGVGVGPG 551
 QY 176 TSAGGL-----LEQAAVEEASDTAAANQ----- 199
 Db 552 VGVGGIPGGVGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 611
 QY 200 LAMNVPAQLAQLAQPTQGT-----TPSSKLGGLWKTVSHRPSINMVEWMANN 247
 Db 612 LGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAG 671


```

DR EMBL; M86355; AAA42271.1; JOINED.
DR EMBL; M86363; AAA42271.1; JOINED.
DR EMBL; M86364; AAA42271.1; JOINED.
DR EMBL; M86366; AAA42271.1; JOINED.
DR EMBL; M86371; AAA42271.1; JOINED.
DR EMBL; M86376; AAA42272.1; JOINED.
DR EMBL; M86373; AAA42272.1; JOINED.
DR EMBL; M86375; AAA42272.1; JOINED.
DR PIR; A36106; EART.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO1500; TROPOELASTIN.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT NON_TER 1
FT SIGNAL <1 21
FT CHAIN 22 864
FT DISULFID 854 859
FT VARSPPLIC 263 307
FT FT
FT FT
FT VARSPPLIC 308 308
FT FT
FT FT
FT FT
FT VARSPPLIC 809 823
FT FT
FT FT
FT SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;

Query Match 7.7%; Score 150; DB 1; Length 864;
Best Local Similarity 25.9%; Pred. No. 0.11; Indels 108; Gaps 18;
Matches 102; Conservative 31; Mismatches 153;

QY 5 GALP-----PEINARMYAGPGSASLVAQAQWSDVSADLSAASAFQSVVWGLTVGSW-- 58
Db 399 GGIPGVGGPGGIGPGVPGCAVSPAAAKAAKAAK--YGARGGVGIPYGVGAGGPPG 456

QY 59 --IGSSAGLMVAASPYVAMSVTAGQALTAQVRVAAAAYETAYGLTVPPPVIAENRA 116
Db 457 YGVGAGAGL-----GGASQAAAA-----AAAKAAKYKAG----- 486

QY 117 ELMILIAITNLIGQNTPAIVNEAEYGENWQAQDAAMFGYAAATATATATALLPPEAPEMT 176
Db 487 -----GAGTIGGLVPG-----AVPGALPGAVPGALPGAVPGALP-GAVPGVP 527

QY 177 SAGLL-----EQAAVEEASDTAAANQLMNVVPAQLQAQPTQTTPSS-KLGLWKTYS 232
Db 528 GTGGVPGAGTPTAAAAAATAKAAK-----KAGQYGLPGVPGVPGVPGVGLPGGVG 579

QY 233 PHRSPISNNVSMANNHSMNTNSGVSM-TNTLSSMLKGF-APAAAQAVQTAQNGVEAMS 290
Db 580 PG-----CVTIGTGTGTGLVPGDLGGAGTPAAKSAKAAKAAKAYRAAA 624

QY 291 SLGSSLGSSGLGGV-----AANLGRAASV---GSLSVPOAWAANQAVTPAARALELT 341
Db 625 GLGAGVPLGVGAGVPGFGAGAGGFGAGAGVPGFGAGAVPGSLAASKAAYGAAGGL--- 681

QY 342 SLTSAERCPGCMGLGLPVGMGARGAG--GGLSGV 374
Db 682 -----CGPGGLGPGGLGPGFGPGGGLGV 708

```

Search completed: November 21, 2003, 16:04:20
Job time : 6.61043 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 70.9578 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-52

Perfect score: 6061

Sequence: 1 MQHHHHHTDVSVGNLEIA.....SAATRRPCRTGRDRWACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6061	100.0	1172	22 AAU01900	M. tuberculosis Tb
2	3866.5	63.8	2502	21 AAB00117	M. tuberculosis po
3	3810	62.9	741	22 AAE05665	Mycobacterium tube
4	2289	37.8	739	16 AAR77501	Malate synthase.
5	2289	37.8	739	22 AAG92334	C glutamicum prote
6	2275	37.5	739	22 AAB79539	Corynebacterium gl
7	2253	37.2	712	22 AAB79540	Corynebacterium gl
8	1619.5	26.7	421	22 AAU08233	Mycobacterium poly
9	1252	20.7	842	15 AAR63567	Feedback inhibito

10	1252	20.7	842	15	AAR63570	Feedback inhibito
11	1248.5	20.6	421	22	AAG64046	Corynebacterium th
12	1248	20.6	842	15	AAR63568	Feedback inhibito
13	1248	20.6	842	15	AAR63569	Feedback inhibito
14	1227.5	20.3	421	21	AAB29607	Modified Coryneb
15	1227.5	20.3	421	24	ABP97764	Amino acid sequenc
16	1224.5	20.2	420	24	ABP97763	Amino acid sequenc
17	1224.5	20.2	421	22	AAU71874	C. glutamicum meta
18	1224.5	20.2	421	22	AAG90027	C. glutamicum prote
19	1224.5	20.2	421	22	AAG79646	Corynebacterium gl
20	1223.5	20.2	421	15	AAR51469	AK alpha subunit T
21	1223.5	20.2	421	15	AAR51465	AK alpha subunit T
22	1223.5	20.2	421	16	AAR71416	C. glutamicum prot
23	1223.5	20.2	421	16	AAR66307	Corynebacterium gl
24	1223.5	20.2	421	19	AAR68147	Aspartokinase alph
25	1223.5	20.2	421	19	AAR69547	Brevibacterium lac
26	1223.5	20.2	421	19	AAW47401	B. lactofermentum
27	1219.5	20.1	421	14	AAR40186	B. flavum AK. Bre
28	1219.5	20.1	421	15	AAR51466	AK wild type alpha
29	1219.5	20.1	421	15	AAR51470	B. lactofermentum
30	1219.5	20.1	421	18	AAW06586	AK wild type alpha
31	1219.5	20.1	421	19	AAW68152	B. lactofermentum
32	1218.5	20.1	421	15	AAR63571	Aspartokinase alph
33	1217.5	20.1	421	22	AAR04857	Feedback inhibito
34	1217.5	20.1	421	23	ABR81982	Corynebacterium gl
35	1215.5	20.1	421	15	AAR63566	C. glutamicum aspa
36	1211.5	20.0	421	22	AAR63566	Feedback inhibito
37	1211.5	20.0	421	22	ABR81989	Corynebacterium sp
38	1176	19.4	505	22	AAU55464	C. glutamicum aspa
39	991.5	16.4	1190	22	ABG18357	Propionibacterium
40	787	13.0	409	21	ABG29516	Novel human diagno
41	785	13.0	411	14	AAR40846	Methylphilus meth
42	773.5	12.8	254	23	ABP65555	Aspartokinase II.
43	770	12.7	156	20	AAY39178	Bifidobacterium lo
44	770	12.7	156	20	ABY39035	M. tuberculosis an
45	741.5	12.2	830	22	AAG25453	M. tuberculosis re
						Novel human diagno

ALIGNMENTS

RESULT 1

AAU01900

ID AAU01900 standard; Protein; 1172 AA:

AC AAU01900;

DT 29-AUG-2001 (first entry)

DE M. tuberculosis TbF14 fusion protein.

XX TBFI4; antigen; vaccine; tuberculosis; AIDS; His tag; MtB81; Mo2;

XX acquired immunodeficiency disease.

XX Synthetic.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

XX Binding-site 3..8

XX /label= Histidine_tag

XX /note= "Nickel chelating region used for purifying

FT the fusion protein"

FT Region 9..749

FT /label= MtB81_region

FT Misc-difference 49

FT /note= "Encoded by TC"

FT Region 750..1172

FT /label= Mo2_region

XX WO200124820-A1.

XX 12-APR-2001.

XX

PF 10-OCT-2000; 2000WO-US28095.
 XX 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 DR N-PSDB; AAS03794.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX Claim 3; Fig 3; 168pp; English.
 XX
 XX The sequence represents Mycobacterium fusion protein antigen Tbf14
 CC consisting of a His tag for purification, antigen MtB81 and antigen Mo2.
 CC Compositions comprising at least 2 heterologous antigens, as a fusion
 CC protein, and vectors expressing the fusion proteins are used as vaccines
 CC to prophylactically immunise mammals (especially humans) against
 CC infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 XX Sequence 1172 AA;
 SQ
 Query Match 100.0%; Score 6061; DB 22; Length 1172;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
 DB 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTQNAL 60
 QY 61 LNARDELQAQIDKWHRRRVEIPIDMAYRQFTEIGYLLPEPDDFTITTSVDABITTA 120
 DB 61 LNARDELQAQIDKWHRRRVEIPIDMAYRQFTEIGYLLPEPDDFTITTSVDABITTA 120
 QY 121 GPOLVPLNARFALNAARWGLSYDALYGTVDVPIPETDGAEGKPTYNKVRGDKVIAYAR 180
 DB 121 GPOLVPLNARFALNAARWGLSYDALYGTVDVPIPETDGAEGKPTYNKVRGDKVIAYAR 180
 QY 181 KFLDDSVPLSSGSGFGDATGFTVQDGLVVALPKSTGLANPGQFAGYTGAESPTSVLLI 240
 DB 181 KFLDDSVPLSSGSGFGDATGFTVQDGLVVALPKSTGLANPGQFAGYTGAESPTSVLLI 240
 QY 241 NHGLHIEILLIDPESQVGTTRDAGVKDVILESATITIMDPEDSVAAVDAADKVLGYRNWLG 300
 DB 241 NHGLHIEILLIDPESQVGTTRDAGVKDVILESATITIMDPEDSVAAVDAADKVLGYRNWLG 300
 QY 301 LNKGDJAAAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHLMINDAIVDT 360
 DB 301 LNKGDJAAAVDKDGTAFRLVNRDRNYTAPGGQFTLPGRSIMFVRNVGHLMINDAIVDT 360
 QY 361 DGSEVFEIGIMDALFTGLIAIHGLKASDVNGPLINSRSGSIYIVKPRMHPAEVATFCLF 420
 DB 361 DGSEVFEIGIMDALFTGLIAIHGLKASDVNGPLINSRSGSIYIVKPRMHPAEVATFCLF 420
 QY 421 SRVEDVLGHPONTMKIGMDEERTTNVNLKACIKAAADRVRVINTGFLDRGTDEHTSWE 480
 DB 421 SRVEDVLGHPONTMKIGMDEERTTNVNLKACIKAAADRVRVINTGFLDRGTDEHTSWE 480
 QY 481 AGPMVRKGTWKSQFWILAYEDHNVDAGLAAGSGRAQVKGWMTMTELMAVMETKIAQP 540
 DB 481 AGPMVRKGTWKSQFWILAYEDHNVDAGLAAGSGRAQVKGWMTMTELMAVMETKIAQP 540
 QY 541 RAGASTAWPSPATAALHALHVOVDVAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
 DB 541 RAGASTAWPSPATAALHALHVOVDVAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600

QY 601 IREEVDNNCSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQALLANLWLRHGV 660
 DB 601 IREEVDNNCSILGYVVRWVDQGVGSKVPDIHDVALMEDRATLRISQALLANLWLRHGV 660
 QY 661 TSADVRLASLERMAPLVDRQNAVGVAYRPMAPNFDSDIAFLAAQELILSGAQQNGYTEPI 720
 DB 661 TSADVRLASLERMAPLVDRQNAVGVAYRPMAPNFDSDIAFLAAQELILSGAQQNGYTEPI 720
 QY 721 LHRRRREFKARAAEKPAAPSDRAGDDAARVQKYGSSVDAERIRRVAAERIVATKQGNV 780
 DB 721 LHRRRREFKARAAEKPAAPSDRAGDDAARVQKYGSSVDAERIRRVAAERIVATKQGNV 780
 QY 781 VVVVSAMGDTTDDLLDAQVCPAPPRELDMLLTAGERISNALVAMATESLGAHARSFT 840
 DB 781 VVVVSAMGDTTDDLLDAQVCPAPPRELDMLLTAGERISNALVAMATESLGAHARSFT 840
 QY 841 GSAQGVITTTGTHGNAKIIDVTPGRLQTALEEGRWLVAGFQGVSDTKDVTTLGRGSDT 900
 DB 841 GSAQGVITTTGTHGNAKIIDVTPGRLQTALEEGRWLVAGFQGVSDTKDVTTLGRGSDT 900
 QY 901 TAVAMAALGADYCEIYTDVVGIFSDADPRIVRNARKLDTVTPEEMLEMAACGAKVLMRC 960
 DB 901 TAVAMAALGADYCEIYTDVVGIFSDADPRIVRNARKLDTVTPEEMLEMAACGAKVLMRC 960
 QY 961 VEYARRHNI PVHVRSSYSRDPGTWVVGSIKDVPMEDPILTGVAHDRSEAKVTIVGLPDIP 1020
 DB 961 VEYARRHNI PVHVRSSYSRDPGTWVVGSIKDVPMEDPILTGVAHDRSEAKVTIVGLPDIP 1020
 QY 1021 GYAAKVPRAVARRRRQHRHGAERLQGRQDRHHLHLIPQTSGLPPPPKWNTRSETRSAS 1080
 DB 1021 GYAAKVPRAVARRRRQHRHGAERLQGRQDRHHLHLIPQTSGLPPPPKWNTRSETRSAS 1080
 QY 1081 TOLLYDDHIKGVSLIGAGVMSHPGVTATFCEALAAVGVNIELISTSEDORSRCCAATPNW 1140
 DB 1081 TOLLYDDHIKGVSLIGAGVMSHPGVTATFCEALAAVGVNIELISTSEDORSRCCAATPNW 1140
 QY 1141 TRPWSRCMKSSGSAATRRPRCTRGRDGRWACQ 1172
 DB 1141 TRPWSRCMKSSGSAATRRPRCTRGRDGRWACQ 1172
 RESULT 2
 AAB00117
 ID AAB00117 standard; Protein; 2502 AA.
 XX AAB00117;
 AC AAB00117;
 XX
 XX 08-FEB-2001 (first entry)
 DT
 XX M. tuberculosis polypeptide sequence comprising Mtb-81 antigen.
 DE
 XX Mycobacterium tuberculosis; antigen; Mtb-91; Mtb-67.2; APC;
 KW antigen presenting cell; serodiagnosis; detection;
 KW human immunodeficiency virus; HIV;
 KW acquired immune deficiency syndrome; AIDS.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 76 /note= "Encoded by TAG stop codon"
 FT
 FT Misc-difference 87 /note= "Encoded by TGA stop codon"
 FT
 FT Misc-difference 88 /note= "Encoded by TAG stop codon"
 FT
 FT Misc-difference 124 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 140 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 174 /note= "Encoded by TAA stop codon"
 FT
 FT Misc-difference 185 /note= "Encoded by TAA stop codon"
 FT

FT Misc-difference 216 /note= "Encoded by TAA stop codon"
FT Misc-difference 217 /note= "Encoded by TGA stop codon"
FT Misc-difference 227 /note= "Encoded by TAG stop codon"
FT Misc-difference 248 /note= "Encoded by TAA stop codon"
FT Misc-difference 259 /note= "Encoded by TAA stop codon"
FT Misc-difference 267 /note= "Encoded by TAA stop codon"
FT Misc-difference 270 /note= "Encoded by TAA stop codon"
FT Misc-difference 305 /note= "Encoded by TAA stop codon"
FT Misc-difference 340 /note= "Encoded by TAA stop codon"
FT Misc-difference 346 /note= "Encoded by TGA stop codon"
FT Misc-difference 360 /note= "Encoded by TAA stop codon"
FT Misc-difference 387 /note= "Encoded by TAG stop codon"
FT Misc-difference 423 /note= "Encoded by TGA stop codon"
FT Misc-difference 436 /note= "Encoded by TGA stop codon"
FT Misc-difference 444 /note= "Encoded by TAA stop codon"
FT Misc-difference 475 /note= "Encoded by TAA stop codon"
FT Misc-difference 492 /note= "Encoded by TGA stop codon"
FT Misc-difference 507 /note= "Encoded by TAG stop codon"
FT Misc-difference 536 /note= "Encoded by TAA stop codon"
FT Misc-difference 550 /note= "Encoded by TAA stop codon"
FT Misc-difference 555 /note= "Encoded by TAG stop codon"
FT Misc-difference 563 /note= "Encoded by TAG stop codon"
FT Misc-difference 572 /note= "Encoded by TAA stop codon"
FT Misc-difference 623 /note= "Encoded by TGA stop codon"
FT Misc-difference 647 /note= "Encoded by TAA stop codon"
FT Misc-difference 750 /note= "Encoded by TAA stop codon"
FT Misc-difference 758 /note= "Encoded by TGA stop codon"
FT Misc-difference 784 /note= "Encoded by TGA stop codon"
FT Misc-difference 816 /note= "Encoded by TGA stop codon"
FT Misc-difference 854 /note= "Encoded by TGA stop codon"
FT Misc-difference 861 /note= "Encoded by TAA stop codon"
FT Misc-difference 917 /note= "Encoded by TGA stop codon"
FT Misc-difference 1035 /note= "Encoded by TAA stop codon"
FT Misc-difference 1082 /note= "Encoded by TAA stop codon"
FT Misc-difference 1160 /note= "Encoded by TGA stop codon"
FT Misc-difference 1161 /note= "Encoded by TAA stop codon"
FT Misc-difference 1164 /note= "Encoded by TGA stop codon"

FT Misc-difference 1187 /note= "Encoded by TAA stop codon"
FT Misc-difference 1188 /note= "Encoded by TAA stop codon"
FT Misc-difference 1211 /note= "Encoded by TGA stop codon"
FT Misc-difference 1253 /note= "Encoded by TGA stop codon"
FT Misc-difference 1344 /note= "Encoded by TGA stop codon"
FT Misc-difference 1348 /note= "Encoded by TAA stop codon"
FT Misc-difference 1386 /note= "Encoded by TAA stop codon"
FT Misc-difference 1527 /note= "Encoded by TAA stop codon"
FT Misc-difference 1579 /note= "Encoded by TAA stop codon"
FT Misc-difference 1582 /note= "Encoded by TGA stop codon"
FT Misc-difference 1616 /note= "Encoded by TGA stop codon"
FT Misc-difference 1628 /note= "Encoded by TGA stop codon"
FT Misc-difference 1643 /note= "Encoded by TGA stop codon"
FT Misc-difference 1673 /note= "Encoded by TGA stop codon"
FT Misc-difference 1683 /note= "Encoded by TAA stop codon"
FT Misc-difference 1685 /note= "Encoded by TAA stop codon"
FT Protein 1691..2438 /label= Mtb-81 polypeptide
FT Misc-difference 2439 /note= "Encoded by TAG stop codon"
FT Misc-difference 2440 /note= "Encoded by TGA stop codon"
FT Misc-difference 2464 /note= "Encoded by TAA stop codon"
FT Misc-difference 2471 /note= "Encoded by TGA stop codon"
FT Misc-difference 2478 /note= "Encoded by TGA stop codon"
FT Misc-difference 2489 /note= "Encoded by TGA stop codon"
FT Misc-difference /note= "Encoded by TAA stop codon"
XX WO200055194-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-US07196.
XX 18-MAR-1999; 99US-0272975.
XX (CORI-) CORIXA CORP.
XX Hendrickson RC, Lodes WJ, Houghton RL;
XX WPI; 2000-638180/61.
XX N-PSDB; AAA53971.
XX Novel Mycobacterium tuberculosis polypeptide comprising an immunogenic
XX portion of M. tuberculosis antigens Mtb-81 and Mtb-67.2, useful for
XX diagnosis, treatment and monitoring therapy of tuberculosis
XX Claim 1; Fig 1a-f; 91pp; English.
XX Polypeptides comprising an immunogenic portion of Mycobacterium
XX tuberculosis antigens Mtb-81 or Mtb-67.2 or variants that differ
XX by substitutions, additions, insertions and/or deletions but which
XX still react with antigen specific antisera or T-cells are described.
CC

CC Also described are polynucleotides encoding the polypeptides. The
CC polypeptides, expression vectors expressing the polypeptides or
CC comprising an antisense polynucleotide, or an antigen presenting
CC cell comprising a sequence encoding the polypeptides are useful for
CC determining the presence or absence of M tuberculosis in whole blood,
CC serum, sputum, plasma, saliva, cerebrospinal fluid or urine in a
CC patient infected with human immunodeficiency virus (HIV).
XX
XX
XX Sequence 2502 AA;
Query Match 63.8%; Score 3866.5; DB 21; Length 2502;
Best Local Similarity 88.3%; Pred. No. 3.8e-286;
Matches 768; Conservative 8; Mismatches 27; Indels 67; Gaps 6;
QY 1 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNAL 60
Db 1691 MQHHHHHTDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNAL 1750
QY 61 LNARDELQAIQDKWHRRIEIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTA 120
Db 1751 LNARDELQAIQDKWHRRIEIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTA 1810
QY 121 GPQLVVPVLNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYAR 180
Db 1811 GPQLVVPVLNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYAR 1870
QY 181 KFLDDSVPLSSGSGFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSILLI 240
Db 1871 KFLDDSVPLSSGSGFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSILLI 1930
QY 241 NHGLHIEILLDPESQVGTTRAGVKDVIIESAITTIMDPEDSVAAVDAADKVLGYRNWLG 300
Db 1931 NHGLHIEILLDPESQVGTTRAGVKDVIIESAITTIMDPEDSVAAVDAADKVLGYRNWLG 1990
QY 301 LNKGDIAAAVDKDGTAFLVLRNDRNYTAPGGQFTLPORSIMFVRNVLMTNDAIVDT 360
Db 1991 LNKGDIAAAVDKDGTAFLVLRNDRNYTAPGGQFTLPGRSLMFRNVLMTNDAIVDT 2050
QY 361 DGSEVFEIGINDALFTGLIAHGIKASDVNGPLNSRTGSIYIVKPKMGHPAEVATCELF 420
Db 2051 DGSEVFEIGINDALFTGLIAHGIKASDVNGPLNSRTGSIYIVKPKMGHPAEVATCELF 2110
QY 421 SRVEDVLGLPONTMKIGIMDEERTTNLKCACIAAADRVFINTGFLDRTGDEIHTSME 480
Db 2111 SRVEDVLGLPONTMKIGIMDEERTTNLKCACIAAADRVFINTGFLDRTGDEIHTSME 2170
QY 481 AGPMVRKGTMSQWILAYEDHNVDAGLAAGFSGRAQVKGKMTTELMDVADVETKIAQP 540
Db 2171 AGPMVRKGTMSQWILAYEDHNVDAGLAAGFSGRAQVKGKMTTELMDVADVETKIAQP 2230
QY 541 RAGASTAWVPSPTAATILHLYHQVDVAAVQOQGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Db 2231 RAGASTAWVPSPTAATILHLYHQVDVAAVQOQGLAGKRRATIEQLLTIPLAKELAWAPDE 2290
QY 601 IREVDNNCOSILGYVVRVWDQVGCSPVDIHDVLMEDRATLRISSQLLANWLPHGVI 660
Db 2291 IREVDNNCOSILGYVVRVWDQVGCSPVDIHDVLMEDRATLRISSQLLANWLPHGVI 2350
QY 661 TSADVRLASLERMAPLVDQVAGVAVRPAPNPFDDSIAPLAAQELILSGAQPNNGYTEPI 720
Db 2351 TSADVRLASLERMAPLVDQVAGVAVRPAPNPFDDSIAPLAAQELILSGAQPNNGYTEPI 2410
QY 721 LHRRRREFKAAAEKAPSDRAGDDAARVQKYGSSVADAEIRRVRAERIVATKQGNVDV 780
Db 2411 LHRRRREFKAAAEKAPSDRAGDDAARVQKYGSSVADAEIRRVRAERIVATKQGNVDV 840
QY 781 VVVVSAMGDTTDDLLDIAQVCPAPPPELMDLLTAGERISNALVAVATIESLCAHARSFT 840
Db 2441 LQIS-----ITLAAARAPPPP-----LRSG-----CXQSPK 2468
QY 841 GSAQGVITTHGNAKIIVDTPGRLQTALE 870
Db 2469 GSXVGC---CHRXA-ITSITPWLXTGLE 2493

RESULT 3
AAE05665
ID AAE05665 standard; peptide; 741 AA.
XX AAE05665;
AC AAE05665;
XX
XX 24-SEP-2001 (first entry)
XX
XX Mycobacterium tuberculosis 88 kDa secreted protein.
XX
XX Mycobacterial disease; infection; 88-kDa protein; tuberculosis; TB;
XX early antigen; human immunodeficiency virus; HIV; secreted protein.
XX
XX Mycobacterium tuberculosis.
XX
XX US6245331-B1.
XX
XX 12-JUN-2001.
XX
XX 31-DEC-1997; 97US-0001984.
XX
XX 02-JAN-1997; 97US-0034003.
XX
XX (UYNV-) UNIV NEW YORK MEDICAL CENT.
XX (COLS) UNIV COLORADO STATE.
XX
XX Laal S, Zolla-Pazner S, Belisile JT;
XX
XX WPI; 2001-424324/45.
XX
XX Detecting a mycobacterial disease (tuberculosis) in individuals
XX comprise assaying a biological sample for the presence of
XX anti-Mycobacterium tuberculosis antibodies or M. tuberculosis
XX antigen-antibody complex
XX
XX Claim 14; Column 121; 96pp; English.
XX
XX The present invention relates to a method for early detection of active
XX mycobacterial disease or infection comprising assaying a biological fluid
XX sample for the presence of early antibodies specific for an 88-kDa
XX Mycobacterium tuberculosis (Mt) protein or immune complexes consisting of
XX an 88-kDa M. tuberculosis protein antigen complexed with an antibody
XX specific for the antigen. The method is useful for the early and rapid
XX detection of mycobacterial disease, particularly tuberculosis, in
XX individuals at heightened risk of developing tuberculosis. This
XX individuals include human immunodeficiency virus (HIV)-infected subjects
XX or other immunocompromised individuals. The method is a rapid and
XX inexpensive screening procedure for detecting mycobacterial disease.
XX The present sequence is a Mt 88 kDa secreted protein early antigen,
XX used in the exemplification of the invention.
XX
XX Sequence 741 AA;

Query Match 62.9%; Score 3810; DB 22; Length 741;
Best Local Similarity 100.0%; Pred. No. 1.3e-282;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNALNARDELQ 68
Db 2 TDRSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNALNARDELQ 61
QY 69 AQIDKWHRRRIEIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTAGPOLVVPV 128
Db 62 AQIDKWHRRRIEIPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAEITTAGPOLVVPV 121
QY 129 LNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYARFLDDSV 188
Db 122 LNARFALNAARNGSLYDALYGTDPETDGAEGKGTYNKVRGDKVIAYARFLDDSV 181
QY 189 LSSGSGFGDATGFTVDGQLVVALPKSTGLANPGQFAGYTGAESPVSILLINHLHIEI 248

Db 182 LSSGSGDGTGFTVQDGLVVALPKSTGLANPGQAGYTGAEBSPTSVLLINHLGHIEI 241
 Qy 249 LIDPESQVGTTRAGVKDVILESALTITIMDFEDSVAADVADKVLGYRNWGLNKGDLAA 308
 Db 242 LIDPESQVGTTRAGVKDVILESALTITIMDFEDSVAADVADKVLGYRNWGLNKGDLAA 301
 Qy 309 AVDKGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHMTNDAIVDTDGEVFEF 368
 Db 302 AVDKGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHMTNDAIVDTDGEVFEF 361
 Qy 369 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVATCELFSEVEDVLG 429
 Db 362 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVATCELFSEVEDVLG 421
 Qy 429 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 488
 Db 422 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 481
 Qy 489 TKKSQWILAYEDHNVDAGLAAGFSGRAOVKGKMWMTLMADWVETKIAOPRAGASTAW 549
 Db 482 TKKSQWILAYEDHNVDAGLAAGFSGRAOVKGKMWMTLMADWVETKIAOPRAGASTAW 541
 Qy 549 VPSPTAATLHALHYHQVDVAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIREEVN 608
 Db 542 VPSPTAATLHALHYHQVDVAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIREEVN 601
 Qy 609 CQSILGYVVRVWDQVGCCKVPDIHDVALMEDRATLRISOLLANWLHGHVITSADYRAS 668
 Db 602 CQSILGYVVRVWDQVGCCKVPDIHDVALMEDRATLRISOLLANWLHGHVITSADYRAS 661
 Qy 669 LERMAPLVDQNAGDVAVRPMAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRREF 728
 Db 662 LERMAPLVDQNAGDVAVRPMAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRREF 721
 Qy 729 KARAAEKAPSDRAGDDAAR 748
 Db 722 KARAAEKAPSDRAGDDAAR 741

RESULT 4

AAAR77501
 ID AAR77501 standard; Protein; 739 AA.
 XX
 AC AAR77501;
 XX
 DT 04-JUN-1996 (first entry)
 XX
 DE Malate synthase.
 XX
 KW aceB gene; malate synthase; Coryneform bacterium; regulatory region;
 XX protein synthesis.
 OS Coryneform glutamicum.
 XX
 PN DE4440118-C1.
 XX
 PD 09-NOV-1995.
 XX
 PF 11-NOV-1994; 94DE-4440119.
 XX
 PR 11-NOV-1994; 94DE-4440119.
 XX
 PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 PI Eikmanns B, Reinscheid D, Sahn H;
 XX
 DR WPI; 1995-374802/49.
 XX
 DR N-PSDB; AAT05501.
 XX
 PT DNA contg. upstream regulatory region from malate synthase gene of
 PT coryneform bacteria - used for regulated synthesis of protein in
 PT coryneform(s), regardless of nutrient medium compen..

PS Example 3; Page 7-8; 12pp; German.

XX AAR77501 is encoded by an approx. 3 kb BfrI-PvuI fragment contg. the
 CC aceB gene from Coryneform glutamicum (see AAT05501). Nucleotides 1
 CC to 574 can be isolated and used to regulate the expression of a
 CC protein-encoding structural gene placed downstream of it, after
 CC incorporation into a vector and transfer into a Coryneform bacterium.
 CC Induction of protein expression occurs even when the culture medium
 CC contains carbon sources other than acetate.

SQ Sequence 739 AA;

Query Match 37.8%; Score 2289; DB 16; Length 739;
 Best Local Similarity 61.3%; Pred. No. 3.6e-166;
 Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

Qy 9 TDRVSVGNRIARVLVDFVNNALPCTDIDPDSFWAGVDKVVADLTTPQNALINARDELQ 69
 Db 20 TERVDAGNQVAKVLYDFVTEAVLPRVGVDAEKFSGFAAARDLTPRNRELLARDELQ 79
 Qy 69 AQIDKWHRRRVIEPIDMDAYROFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 128
 Db 80 MLIDDDYHRNN-SGTIDQEAYEDFLKEIGYLVPEPEAAEIRTQNVDTIEISSTAGPQLVVPV 138
 Qy 129 LNARFALNANARWGLSYDALYDGTVDIPETDGAEGKFTYKVRGDKVIAYARKFLDDSV 188
 Db 139 LNARFALNANARWGLSYDALYGTNAIPETDGAEGKFTYKVRGDKVIAYARKFLDDSV 198
 Qy 189 LSSGSGDGTGFTVQDGLVVALPKSTGLANPGQAGYTGAEBSPTSVLLINHLGHIEI 248
 Db 199 LDGASHADVEKNITDGLAAHIGDSVYRLKRESYRGFTGNFLDPEAILLETNGHIL 258
 Qy 249 LIDPESQVGTTRAGVKDVILESALTITIMDFEDSVAADVADKVLGYRNWGLNKGDLAA 308
 Db 259 QIDPVHPIGKADTKGLKDIVLESATITIMDFEDSVAADVADKVLGYRNWGLNKGDLAA 318
 Qy 309 AVDKGTAFRLVLRNDRNTAPGGQFTLPGRSLMFVRNVGHMTNDAIVDTDGEVFEF 368
 Db 319 EMSKNGRIFTRELNDKRVYIGRNGTELVLHGRSLLFVRNVGHLMQNPSIL-IDGEEIFEG 377
 Qy 369 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVATCELFSEVEDVLG 428
 Db 378 IMDALFTGLIAHLKASDVNGPLINSRTGSIYIVKPKHGPAPAEVATCELFSEVEDVLG 434
 Qy 429 LPQNTMKIGIMDEERTITVNLKACIKAAADRVWFINTGFLDRTGDEIHTSMEAGPMVRKG 488
 Db 435 LPRHTLKVGWMBEERTSVNLDASIMEVADRLAFINTGFLDRTGDEIHTSMEAGPMVRKA 494
 Qy 489 TKKSQWILAYEDHNVDAGLAAGFSGRAOVKGKMWMTLMADWVETKIAOPRAGASTAW 548
 Db 495 DMQTAPWKQAYENNVNDAGIQRGLPQAGIKGGMWMTLMADWVETKIAOPRAGASTAW 554
 Qy 549 VPSPTAATLHALHYHQVDVAVQOGL-AGKRRATIEQLLTIPLAKELAWAPDEIREEVN 607
 Db 555 VPSPTGATLHATHYHLVDVFKVQDELRAAGRDRSLRNILTIPTAPNTNWSSEKKEMDN 614
 Qy 608 NCOSILGYVVRVWDQVGCCKVPDIHDVALMEDRATLRISOLLANWLHGHVITSADYRAS 667
 Db 615 NCOSILGYVVRVWEHGVGCCKVPDIHDIDIMEDETRATLRISOLLANWLHGHVITSADYRAS 674
 Qy 668 SLERMAPLVDQNAGDVAVRPMAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRREF 727
 Db 675 SLERMAPLVDQNAGDVAVRPMAPNFDSDIAFLAAQELILSGAQQPNGYTEPILHRRREF 734
 Qy 728 FKAR 731
 Db 735 FKAR 738

RESULT 5
 AAG93234
 ID AAG93234 standard; Protein; 739 AA.
 XX

AC AAG93234;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6988.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 KW Corynebacterium glutamicum.
 OS
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH68453.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 29; SEQ ID NO: 6988; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 739 AA;
 Query Match 37.8%; Score 2289; DB 22; Length 739;
 Best Local Similarity 61.3%; Pred. No. 3.6e-166;
 Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;
 QY 9 TDRVSVGNLRIARVLYDFVNEALPGTDIDPSDFWAGVDKVVADLTPQNCALLNARDELQ 68
 DB 20 TERVDAGGMQVAKVLYDFVTEAVLPRGVDAEFNSGFPAIARDLTPRRELLARDELQ 79
 QY 69 AQIDKWHRRRVIPIIDMAYRQFLTEIGYLLPEPDDFTTTSGVDAEITTTAGPQLVWPV 128
 DB 80 MLIDDVHRNN-SGTIDQEAIVEDFLKEIGYLVPEEAAEIRTONVDTEISSTAGPQLVWPI 138
 QY 129 LNAEFALNANARWSLYDALYCTDVIPTDGAEGKFTYKRGDKVIAYAKSEFLDDSV 188
 DB 139 LNAEFALNANARWSLYDALYCTNAPETDGAEGKFTYKRGDKVIAYAKSEFLDDSV 198
 QY 189 LSSGSFGDAGFTVQDQQLVVALPDKSTGLANPGQFAGYTGAAESPTSVLLINHLHIEI 248
 DB 199 LDGASHADVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIEL 258
 QY 249 LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADKVLGYRNWGLNKGDLAA 308

DB 259 QIDPVHPGKADTKGLKDIVLESATITIMDFEDSVAADKVTLCYSNFWGLNTGELKE 318
 QY 309 AVDKDGTAFRLVLRDRNYTAPGGQFTLPGRSLMFVRNVGHMTNDAIVDTDGSSEVFG 368
 DB 319 EMSXNGRIFTRLNKORVYIGRNGTBLVHGRSLLFVRNVGHLMQNPISL-IDGEEIFEG 377
 QY 369 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKMHGFAEVAFTCELSRVEDVLG 428
 DB 378 IMDAVLTTCVCAIPGAPQN---KWRNSRKSGSIYIVKPKHGPPEVAFTNELFCRVEDLLD 434
 QY 429 LPONTMKGIMDEBERRITVNLKACIKAADRVVFIINTGLDRTGDEIHHSMEAGPNVRKG 488
 DB 435 LPRHTLKVGVNDEBERRITVNLKASIMEVADRLAFINTGLDRTGDEIHHSMEAGPNVRKA 494
 QY 489 TMRKSPWILAYEDHNVDAGLAAGFSGRAQVGVKGMWMTMTLMADMVETKIAQPRAGASTAW 548
 DB 495 DMQTAPWKQAYENNVVDAGIQRGLPGKAQIGKGMWMTMTLMADMVETKIAQPRAGASTAW 554
 QY 549 VPSPTAATLHALYHQVDVAAVQOGL-AGKRRATISQLLTIPLAKELAWADPDIREEVDN 607
 DB 555 VPSPTGATLHATHYHLVDVEKVDDELRAAGRDSLRNLTITPTAPNTNWSSEKKEEMDN 614
 QY 608 NCQSILGVVVRWVDQGVGCSKVPDIHDVLAWEDETRATLRISQALLANLWLRHGVITSADVRA 667
 DB 615 NCQSILGVVVRWVEHGVGCSKVPDIHDIDNEDETRATLRISQALLANLWLRHGVITSADVRA 674
 QY 668 SLERMAPLVDRQAGDVAYRPMAPNFDSDSIAFLAAQELILSGAQQPNVGYTEPILHRRRE 727
 DB 675 SLERMAVVVDKQAGDEAYRDMAPNYDASLAFQAAXDLIFEGTKSPSGYTEPILHRRRE 734
 QY 728 FKAR 731
 DB 735 FKAR 738
 RESULT 6
 AAB79539
 ID AAB79539 standard; Protein; 739 AA.
 AC AAB79539;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:594.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 OS
 XX
 PN WO200100844-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00943.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.

```

PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99DE-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
XX N-PSDB; AAF71656.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 20; Page 992-995; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 739 AA;

Query Match 37.5%; Score 2275; DB 22; Length 739;
Best Local Similarity 61.0%; Pred. No. 4.2e-165;
Matches 442; Conservative 90; Mismatches 186; Indels 6; Gaps 4;

QY 9 TDRVSGVNLRTARVLVDYFNNEALPGCTIDDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 20 TERVDAGGVQVAKVLDYFVTEAVLPRVGVDAEKFWSGFAATARDLTPNRELLARRDELQ 79
QY 69 AQIDKWHRRVIERIDMAYQFTEIGYLLPEPDDFTITTSVGVDAEITTTAGPQLVVPV 128
DB 80 MLIDDYHNN-SGTDQAEYEDFLKEIGLYVEEPEAAIRTONVDTEISSTAGPQLVVPV 138
QY 129 LNARFALNANRWGSLDYALGTVDVIBETDGAEGKPTYNKVRGDKVIYARKFLDSDVP 188
DB 139 LNARFALNANRWGSLDYALGTVNAIBETDGAEGKPTYNKVRGDKVIYARKFLDSDVP 198
QY 199 LSSGSFGATGTFTVDGOLVWALPKSTGLANPGQFAGYTGABSPSTVLLINHLHTEI 248

```

```

DB 199 LDGASHADVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIEL 258
QY 249 LIDPESQVGTTRAGVKOVILESAITTIMDFDSVAADAAKVLGYRWLGNKCDLAA 308
DB 259 QIDPVHPIGKADTKGLKDIVLESAITTIMDFDSVAADAAEDKTIQYSWNFGTNGELKE 318
QY 309 AVDKGTAFRLVNRDRNYTAPGGQOFTLPGSLMFVRNVGHMLTNDATVDTDGSEVFEG 368
DB 319 ESKNGRIETRELKNDRVYIGRNGTELVLRGSLLFVRNVGHLMQNPISL-IDGEEIFEG 377
QY 369 INDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVLG 428
DB 378 INDVLTTCALPGIAPQN---KMNRSRKGSIYIVKPKMHGPAEVAFTCELSRVEDVLG 434
QY 429 LPQNTMKIGIMDEERTTNLACIAAADRVRVFTNGTLDRTGDEIHTSMGAPMVRKG 488
DB 435 LPRHTLVKGVMBEERTSVNLDASIMEVADRLAFINTGELDRGTGDEIHTSMGAPMVRKA 494
QY 489 TMKSQWILAYEDHNVDAAGAFSGRAQVKGKMTMTLMADMYETKTAOPRAGASTAW 548
DB 495 DMQTAPWKQAYENNVDAGIQRLPGKAIQKGMAMVTELMAMENLEKKIGQREGANTAW 554
QY 549 VPSPTAATLHALHYHOVDVAAVQOGL-AGKRRTATIEQLLTIPLAKELAWAPDSIREVDN 607
DB 555 VPSPTGATLHATHYHLVDYFKVQDELRAAGRRDLSRLNLTISTAPNTNWSBEKKEDMN 614
QY 608 NCOSILGYVRVVDQVGCCKVPDIHDVALMEDRATLRISQQLLANLWLRHGVITSADVRA 667
DB 615 NCOSILGYVRVWVHGVGCCKVPDIHDIDLMEDRATLRISQQLANLWLRHGVITSADVRA 674
QY 668 SLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELJLSGAQQPENGYTEPILHRRRE 727
DB 675 SLERMAVVDKQAGDEAYDAPKDYDASLATQAAKDLIFEGTKSPSGVTEPILHARRRE 734
QY 728 FXAK 731
DB 735 FXAK 738

RESULT 7
AAB79540
ID AAB79540 standard; Protein; 712 AA.
XX
AC AAB79540;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:596.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.

```

PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1032180.
 PR 08-JUL-1999; 99DE-1032227.
 PR 08-JUL-1999; 99DE-1032230.
 PR 08-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX WPI; 2001-061975/07.
 DR N-PSDB; AAF71657.
 DR
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 998-1000; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. Glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. Glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 712 AA;
 Query Match 37.2%; Score 2253; DB 22; Length 712;
 Best Local Similarity 61.2%; Pred. No. 1.9e-163;
 Matches 438; Conservative 89; Mismatches 183; Indels 6; Gaps 4;
 QY 17 LRIARVLYDFVNEALPGTIDPDSFWAGVKVADLTTPQNALNARDELQADKWHR 76
 Db 1 MQVAKVLYDFVTAVLPVGVDAEKFSWGFALARDLTPRNRELLARRDELQMLDDYHR 60
 QY 77 RRVIEPIDMDAYRQFUTEIGYLLPEPDDFTITTSQVDAEITTAGPQLVVPVYNARFALN 136
 Db 61 NN-SGTIDQEAEDFLKEIGYLVVEEPAEAINTQNVDTISISSTAGPQLVVPVYNARFALN 119
 QY 137 ANARWGSYDALYGDVTPETDGAERKGYTNKVGDKVIAYARKFLDSDVPLSGSGCD 196
 Db 120 ANARWGSYDALYGINAIPETDGAERKGYTNKVGDKVIAYARKFLDSDVPLSGSGCD 179

QY 197 ATGFTVQDQGLVVALPKSTGLANPCQFAGYTGAAARSPTSVLLINHLHIEILLDESQV 256
 Db 180 VEKYNITDGLAAHAGDSVYRLKXNRSYRGFTNFDPEAILLETNGLHIEQLDPVHP 239
 QY 257 GTTDRAGVKDVILESATITIMDPEDSVAAVDAADKVLGYRNWGLNKGDLAAAVDKDGA 316
 Db 240 GKADKTGLKDIVLESATITIMDPEDSVAAVDAADKLTGYSNWFGLNTGELKEEWSXNGRI 299
 QY 317 FLRVINDRNYTAPGGQFTLPGRSLMFVNVNCHLMTNDAIVDTDGVSEFEGIMDALFTG 376
 Db 300 FTRELNKDRVYIGRNGTELVHGRSLLFVNVNCHLMTNDAIVDTDGVSEFEGIMDALFTG 358
 QY 377 LIAIHGLKASDVNGPLNSRTGSIYIVKPKWHPAEVAFCELFPSRVEDVLGPQNTMKI 436
 Db 359 VCALPGIAPQV---KWRNSRKGSIYIVKPKQHPFEVAFNLFGRVEDLLDPRHTLKV 415
 QY 437 GIMDEERTTNLAKACIKAAAADRVFINTGFLDRTGDEIHTSMAGPMVRKGTMSQPMI 496
 Db 416 GWDEERTSVNLDAIMEVADRLAFINTGFLDRTGDEIHTSMAGPMVRKGTMSQPMI 475
 QY 497 LAYEDHNVDAGLAAGFSGRAOVGKGMWMTTELKADWVETKIAOPRAGASTAWPSPSTAAT 556
 Db 476 QAYENNVDAGIQRGGLPKAQIQKGMWMTTELKADWVETKIAOPRAGASTAWPSPSTAAT 535
 QY 557 LHALHYHQVDVAAVQOGL-AGKERATIEQLLTITLAKELAWAPDEIRREEVDNNCQSILGY 615
 Db 536 LHATHYHLVDVFKVQDELRAAGRRDSLRNLTITSTAPNTNWSSEKKEEDMNCQSILGY 595
 QY 616 VVRVVDGVCCKVPDIHDVVALMEDRATLRISSQLLANWLHGVITTSADVRSLEMAPL 675
 Db 596 VWRVHEGVGCKVPDIHDVVALMEDRATLRISSQLLANWLHGVITTSADVRSLEMAPL 655
 QY 676 VDRQAGDVAYRPMAPNFDSDIAFLAAQELILSGAOPNGYTPILHRRRREFKAR 731
 Db 556 VDRQAGDEAYRPMAPNFDSDIAFLAAQELILSGAOPNGYTPILHRRRREFKAR 711
 RESULT 8
 AAU08233
 ID AAU08233 standard; Protein; 421 AA.
 AC AAU08233;
 DT 17-DEC-2001 (first entry)
 DE Mycobacterium polypeptide for MO-2.
 KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
 immunostimulant; MO-2.
 OS Mycobacterium sp.
 FH Key Location/Qualifiers
 FT Misc-difference 41 /note= "Encoded by T"
 FT Misc-difference 305 /note= "Encoded by GAC"
 FI
 FN WO200162893-A2.
 XX
 PD 30-AUG-2001.
 PF 26-FEB-2001; 2001WO-US05992.
 XX
 PR 25-FEB-2000; 2000US-0185037.
 PR 08-AUG-2000; 2000US-0223828.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
 XX WPI; 2001-536638/59.
 DR

DR N-PSDB; AAS12505.

XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from

PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune

PT response to and inhibiting development of a Mycobacterium infection -

XX

PS Claim 1; Fig 17; 161pp; English.

XX

CC The present invention relates to the isolation of Mycobacterium

CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids

CC encoding them. The invention describes compounds and methods for the

CC diagnosis of tuberculosis or for inducing protective immunity against

CC tuberculosis. The compounds comprise at least one immunogenic portion

CC of one or more Mycobacterium proteins and nucleic acid molecules

CC encoding such polypeptides. The Mycobacterium proteins and nucleic acid

CC molecules encoding them can be used in diagnostic kits for the detection

CC of Mycobacterium infection in patients and biological samples. The

CC compounds of the invention and antibodies directed against the

CC Mycobacterium proteins may be used in vaccines for immunisation against

CC Mycobacterium infections. The nucleic acids encoding the Mycobacterium

CC proteins may be used in gene therapy. The present sequence represents

CC Mycobacterium polypeptide for MO-2.

XX

SQ Sequence 421 AA;

Query Match 26.7%; Score 1619.5; DB 22; Length 421;

Best Local Similarity 85.1%; Pred. No. 2.7e-115;

Matches 338; Conservative 8; Mismatches 34; Indels 17; Gaps 3;

QY 749 VQKYGSSVDAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLLDLAQVCPAPPPR 808

DB 5 VQKYGSSVDAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLLDLAQVCPAPPPR 64

QY 809 ELDMLLTAGERISNALVAMAEIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPEGLQTA 868

DB 65 ELDMLLTAGERISNALVAMAEIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVTPEGLQTA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVMAAALGADVCEIYTDVGDGIFSAADP 928

DB 125 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVMAAALGADVCEIYTDVGDGIFSAADP 184

QY 929 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPTGVVVG 988

DB 195 RIVRNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPTGVVVG 244

QY 989 IKDVPMEPIILTVGAHDRSEAKVTIVGLPIGVAAKVFRAVARRRQRHGAEE----- 1043

DB 245 IKDVPMEPIILTVGAHDRSEAKVTIVGLPIGVAAKVFRAVA-----DADVNIDM 295

QY 1044 PLOGRGRQDRHLLLPQTS---GPPPKWTRSTRSASTOLLYDDHIGKVSILIGAMR 1100

DB 236 VLQNVSKVEKGTDTFTCSRVDGPAAVEKLDLSLRNEIGFSQLLYDDHIGKVSILIGAMR 355

QY 1101 SHPGVTATFCBALAAGVGNIELISTSEDPQRSCCAAT 1137

DB 356 SHPGVTATFCBALAAGVGNIELISTSEIRISVLCRDT 392

RESULT 9

AAAR63567

ID AAR63567 standard; Protein; 842 AA.

XX

AC AAR63567;

XX

DT 29-JUN-1995 (first entry)

XX

DE Feedback inhibition release aspartokinase 279Aa, 301ser, 308Thr.

XX

KW Feedback inhibition release aspartokinase; Brevibacterium flavum;

KW Corynebacterium; L-lysine production.

XX

OS Brevibacterium flavum.

XX

PN JP06261766-A.

XX

PD 20-SEP-1994.

XX

PF 16-MAR-1993; 93JP-0055451.

XX

PR 16-MAR-1993; 93JP-0055451.

XX

PA (MTP) MITSUBISHI PETROCHEMICAL CO LTD.

XX

DR WPI; 1994-337417/42.

DR N-PSDB; AAQ72693.

XX

PT DNA encoding feedback inhibition release asparto-kinase - used

PT for the production of L-lysine.

XX

PS Claim 4; p16; 28pp; Japanese.

XX

CC AAQ72693 encodes AAR63567 Brevibacterium flavum feedback inhibition

CC release aspartokinase, it was used in the construction of a

CC plasmid. The plasmid was used to transform Corynebacterium cells,

CC where the expression of AAQ72693 resulted in increased L-lysine

CC production.

XX

SQ Sequence 842 AA;

Query Match 20.7%; Score 1252; DB 15; Length 842;

Best Local Similarity 36.9%; Pred. No. 1e-86;

Matches 346; Conservative 128; Mismatches 240; Indels 224; Gaps 31;

QY 255 QVSTTRAGVKDVIKESAITTMDPDSVAADAAKVLGYRNWLGKNGDLAAAVDKDG 314

DB 25 RIVATKKAGNVVVCAMG-----DTTDELL-----ELAAAVNP-- 59

QY 315 TAFRLVLRDRNVTAGGGQFTLPGRSLMFVRNVGHMTNDALVDTGDSSEVPEGIMDALP 374

DB 60 -----VPPAEEMDLITAG-----ERISNALV 81

QY 375 TGLIAIHGLKASDVNGPLINSRTGSIYIVKPMKHGPAEVAFTCELSRVEDVL----- 427

DB 82 AMAIESLGAESAQSFTG---SQAG---VLTTRHGNARIVDVTP---GRVREALDEKICI 132

QY 428 -----GLPQNTMKIGIMDE--ERRITVNLKACIKA-----AADRVVFINTGF 467

DB 133 VAGFGQVKNRETDRVTTLGRGSDTTAVALLAALNADVCEIYSDVDGVYTDPRIVENAQK 192

QY 468 LDRGTGDEIHTSMEA-GPMVRKGTMKSQPWILAY-----EDHNVDAAG-LAAGESGRAOV 518

DB 193 LEKLSFEEMLELAAGVSKIL--VLRSEYARAFENVPLRVSSYSNDPGTLIAGSMEDIPV 250

QY 519 GKGMWTMTMLADMVETKIAQPRAGASTAWPSPATAATHALHYHQVDVAAVQQGLAGKR 578

DB 251 EEAV--LTGVATDKSEAKVTV--LGISDK--PGEAAKVFPRALADABINIDMWLVQNVSV 304

QY 579 RATIEQLLTIP-----LAKELAWAPDEIRREEDNNQCSTLGVVVRWVQ----- 622

DB 305 DGTDTITFTCPRSDGRRAMEILKKL-----VQGNWTVNL-----YDQGVKVSIV 350

QY 623 GVGCSKVPDIHDVALMEDRATLRISQLLANMLRHGVITTSADVRSLESLRMAPLVDRQNA 682

DB 351 GAGMKSHPGV-TAEFNEALRDVNVNIEL-----ISTSEIRIS-----VLIREDDL 394

QY 683 DVAYRPMANFDDSIAPLAAQELILSGAQQNGYTEPILHRRRRERFKARAAEKAPASDRA 742

DB 395 DAAARALHEQF-----QLGGEDEAVVYA-----GT 419

QY 743 GDDAARVQKVGSSVDAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLLDLAQVCP 802

DB 420 GRVALVQKVGSSVDAEIRIRVAERIVATKQGNVWVVSAMGDTTDELELLELAAYN 479

QY 803 PAPPRELMDMLTAGERISNALVAMAEIESLGAHARSFTGSGAGVITTTGTHGNAKIIDVT 862

DB 480 FVPPAREMDMLTAGERISNALVAMAEIESLGAESAQSFTGSGAGVLTTRHGNARIVDVTP 539


```
XX JN2001120270-A.
XX PD 08-MAY-2001.
XX PF 01-NOV-1999; 99JP-03111148.
XX PR 01-NOV-1999; 99JP-03111148.
XX PA (AJIN ) AJINOMOTO KK.
XX DR WPI: 2001-364760/38.
XX DR N-PSDB; AAH45374.
XX
PT A heat-resistant lysin biosynthetic system enzyme gene of a high
PT temperature-resistant coryneform microbe
XX Claim 6; Page 21-22; 27pp; Japanese.
XX
CC The invention relates to a gene from a high temperature-resistant
CC coryneform microbe that encodes a heat-resistant lysin biosynthetic
CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity
CC and can be used for growing amino acid-producing microbes. The
CC present amino acid sequence corresponds to an enzyme of the
CC invention.
XX
XX Sequence 421 AA;
SQ Query Match 20.6%; Score 1248.5; DB 22; Length 421;
Best Local Similarity 65.2%; Pred. No. 6.9e-87;
Matches 251; Conservative 58; Mismatches 63; Indels 13; Gaps 3;
QY 749 VQYGGSSVADAEIRRVARIVATKKGNDVVVVVVSAMGDTTDLDLAQVCPAPPFR 808
DB 5 VQYGGSSLESAERIRNVAERIVATKAGNDVVVVVCSAMGDTTDLDLAAAVNPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMATESLCAHARSFTGSOAGVITTTGHNAKIIDVTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVAMATESLGAESAQSTGSGAGVLTTERHGNARIVDVTFRVREA 124
QY 869 LEEGRVILVAGQGVSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFISADP 928
DB 125 LDEGKICIVAGQGVNKTETDVTTLGRGSGDTTAVALLAALGADVCEIYSDVDGVYADP 184
QY 929 RIVRNARKLDVTFEEMLEMAACGAKVLMRCVEYARHNIPVHVRSSYSRPGTGVVGS 988
DB 185 RIVPNAQKLERISFEEMLELAAGVSKILVLRSVEYARAFNFMVRSSYSNDPGTLIAGS 244
QY 989 IKDVPMEDEPILTVGAHDSKAKVTIVGLPIFGYAAKVFRAVARR-----RQHRHGAA 1042
DB 245 MEDIPMEEAVALTVGATDKSEAKVTIVGLPIDKFGEBAKVFRALADAENIDMVLQNVSSVE 304
QY 1043 ERLQGRGRQDRHLLHLLPQTSGPPPWKNWTRSETASSTQLLYDDHIGKVLGAGMRSH 1102
DB 305 D-----GTTD--ITFTCPRSDGPRAMELLKKVQQQGDWNTNLYDDQGVKSLVAGMKSH 357
QY 1103 PGVATTFCEALAAVGVNTELISTSE 1127
DB 358 PGVTAEPMEALRDVNVNELISTSE 382
RESULT 12
ID AAR63568
XX AAR63568 standard; Protein; 842 AA.
XX
AC AAR63568;
XX
XX 29-JUN-1995 (first entry)
XX
DE Feedback inhibition release aspartokinase 279Thr, 301Ser, 308Thr.
XX
KW Feedback inhibition release aspartokinase; Brevibacterium flavum;
KW Corynebacterium; L-lysine production.
```

```
XX Brevibacterium flavum.
XX
XX JN2001120270-A.
XX PD 08-MAY-2001.
XX PF 01-NOV-1999; 99JP-03111148.
XX PR 01-NOV-1999; 99JP-03111148.
XX PA (AJIN ) AJINOMOTO KK.
XX DR WPI: 2001-364760/38.
XX DR N-PSDB; AAH45374.
XX
PT A heat-resistant lysin biosynthetic system enzyme gene of a high
PT temperature-resistant coryneform microbe
XX Claim 6; Page 21-22; 27pp; Japanese.
XX
CC The invention relates to a gene from a high temperature-resistant
CC coryneform microbe that encodes a heat-resistant lysin biosynthetic
CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity
CC and can be used for growing amino acid-producing microbes. The
CC present amino acid sequence corresponds to an enzyme of the
CC invention.
XX
XX Sequence 842 AA;
SQ Query Match 20.6%; Score 1248; DB 15; Length 842;
Best Local Similarity 36.8%; Pred. No. 2.1e-86;
Matches 345; Conservative 128; Mismatches 241; Indels 224; Gaps 31;
QY 255 QVGTTRAGVKVILESAITTIMDFSDVAADVDAADKVLGYRNWLGKGLAAAVDKDG 314
DB 25 RIVATKKAGNVVWCSAMG-----DTTDELL-----ELAAAVNP-- 59
QY 315 TAFRLVLRDRNTAPGGQFTLPGBSLMFVRNVGHMLNDALVDGSEVFEGIMDALF 374
DB 60 -----VPPAEEMDLMTAG-----ERISNALV 81
QY 375 TGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAEVAFTCELSRVEDVL----- 427
DB 82 AMAESLGABEAQSFSG-----SQAG---VLTERHGNARIVDVTTP--GRVREALDEGKICI 132
QY 428 -----GLPQNTWKIGIMDE--ERRTVNLKACIK-----AADRVVVFINTGF 467
DB 133 VAGFQGVNKTETDVTTLGRGSGDTTAVALLAALNADVCEIYSDVDGVYADPPIVNAQK 192
QY 468 LDRGTDEIHTSMEA-GPMVVRKGTMSQPMILAY-----EDHNVDAAG-LAAGFSGRAQV 518
DB 193 LEKLSFEEMLELAAGVSKIL--VLRSEYARAFNVLPRVSSYSNDPGTLIAGSMEDIPV 250
QY 519 GKGHWMTMELMADVMETKIAQPRAGASTAWPPSPATAIHALHYQVDVAVAOQGLAGKR 578
DB 251 EEAV--LTGVATDKSEAKVTV--LGI-SDK--PGEVAKVFRALADAENIDMVLQNVSSVE 304
QY 579 RATIEQLLTIP-----LAKELAWAPDEIEEYDNNNCOSILGYVVRVDO----- 622
DB 305 DGTDTITFTCPRSDGRRAMEILKKL-----QVGGNWTNVL-----YDDQGVKSVLV 350
QY 623 GVGCSKVPDIHDVALMEDRATILFRISQLLANMLRHGVITTSADVRSILRMAPLVDQRNAG 682
DB 351 GAGWKSHPGV-TAEFMEALRDVNVNIEL-----ISTSEIRIS-----VLIREDDL 394
QY 683 DVAYRPMANFNDDSIAPLAAQELILSGAQPGNYGTEPILHRRRRREFKARAAEKAPASDRA 742
DB 395 DAAARALHEQF-----QLGGEDEAVVYA-----GT 419
QY 743 GDDAARVQKYGGSSVADAERIRRVARIVATKKGNDVVVVVVSAMGDTTDLDLAQVVC 802
DB 420 GRVALVQKYGGSSLESABEIRNVARIVATKAGNVVWVCSAMGDTTDLLELAANV 479
```


KW L-lysine; L-threonine; insensitive; product inhibition;
 KW coryneform bacterium; mutant; mutain.
 OS Corynebacterium glutamicum.
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 311 /note= "Wild-type Thr replaced with Ile"
 FT
 PN WO200063388-A1.
 XX
 XX 26-OCT-2000.
 XX
 XX 14-APR-2000; 2000WO-JP02456.
 PF
 XX 19-APR-1999; 99JP-0110437.
 PR
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Yokoi H, Ohnishi J, Ochiai K, Yonetani Y, Ozaki A;
 PI WPI; 2000-687179/67.
 XX DR N-PSDB; AAC64119.
 DR
 XX DNA encoding modified aspartokinase without synergistic feedback
 PT inhibition by L-lysine and L-threonine for efficient production of
 PT L-lysine by coryneform bacterium fermentation -
 XX
 XX Claim 1; Page 29-30; 33pp; Japanese.
 PS
 XX The invention relates to a modified aspartokinase from Corynebacterium
 CC glutamicum (e.g., AAB29607) and to DNA encoding the modified
 CC aspartokinase (e.g., AAC64119). The aspartokinase of the invention is
 CC modified to remove its sensitivity to feedback inhibition by L-threonine
 CC or L-lysine. This is achieved by substituting the wild-type Thr 311 with
 CC any other amino acid; for example, in AAB29607 Thr 311 is substituted by
 CC Ile. The invention also relates to expression vectors and host
 CC coryneform bacteria containing the modified DNA, and to the preparation
 CC of L-lysine using coryneform bacteria transformed with the modified
 CC aspartokinase-encoding DNA. Culturing coryneform bacteria transformed
 CC with the modified aspartokinase DNA of the invention provides efficient
 CC production of L-lysine. The present sequence represents a modified
 CC Corynebacterium glutamicum aspartokinase of the invention, T3111.
 XX
 SQ Sequence 421 AA;
 Query Match 20.3%; Score 1227.5; DB 21; Length 421;
 Best Local Similarity 64.0%; Pred. No. 2.8e-85;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;
 QY 749 VQYGGSSVADAEIRRVAAERIVATKKQGNVWVVSAMGDTTDLDLAQVCPAPPPR 808
 Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVCSAMGDTTDLLELAANVPVPPAR 64
 QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLQTA 868
 Db 65 EMDMLLTAGERISNALVAMAIESLGAESAQFTGSQAGVLTTERHGNARIVDVTGVRVREA 124
 QY 869 LREGSVLVAGQGVSDTKVTTLGRGSDTAVAMAALGADYCEIVTDVDFGFSADP 928
 Db 125 LDEGKICIVAGQGVNKEIRDTVTLGRGSDTITVALAALNADYCEIYSDVDGVTADP 184
 QY 929 RIVRNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNIPVHVRSYSDRPFTVWVGS 988
 Db 185 RIVPNAQKLEKLSFEEMLEAAVGSKILVLRSEYARAFNVLVRSSYSDNDPGTLIAGS 244
 QY 989 IKDVNEMDPIILGVADHREAKVTIVGLPDIPGYAAKVFRAV-----RRRQHR 1038
 Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGISDKPGEAAKVFRAADAEINIDVMLQNVSVE 304
 QY 1039 HGAERLQGRQD-RHHLHLLP--QTSGPPPPKWNTRSETRSASTQLLYDDHIGKVSLLI 1095
 Db 65 EMDMLLTAGERISNALVAMAIESLGAESAQFTGSQAGVLTTERHGNARIVDVTGVRVREA 124

Db 305 DGTDTIIFTCPRSDGRRAMEILKKLVQVG-----NM-----TNVLVDDQGVKSLV 350
 QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGMKSHPGVTAFEMEALRDVNVNIELISTSE 382
 RESULT 15
 ID ABP97764 standard; Protein; 421 AA.
 AC ABP97764;
 XX
 XX 28-MAY-2003 (first entry)
 DT
 XX Amino acid sequence of a aspartate kinase carrying T3111 mutation.
 DE
 XX Coryneform bacteria; chemical compound; L-amino acid; vitamin; lysC;
 KW aspartate kinase.
 KW
 XX Synthetic.
 OS Corynebacterium glutamicum.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Met encoded by GTG"
 FT Misc-difference 311 /note= "wild type Thr changed to Ile"
 FT
 XX WO2003014330-A2.
 PN
 XX 20-FEB-2003.
 PD
 XX 30-JUL-2002; 2002WO-EP08465.
 PP
 XX 06-AUG-2001; 2001US-309877P.
 PR
 XX (DEGS) DEGUSSA AG.
 PA
 XX Bathe B, Kreutzer C, Moeckel B, Thierbach G;
 PI WPI; 2003-278476/27.
 XX DR N-PSDB; ABZ77392.
 DR
 XX Coryneform bacteria to produce chemical compounds e.g. L-lysine, has at
 PT least two copies of open reading frame, gene or allele in question at a
 PT particular desired site -
 FT
 XX Claim 16; Page 105-106; 109pp; English.
 PS
 XX The specification describes Coryneform bacteria which produce chemical
 CC compounds, which instead of a single copy of an open reading frame (ORF),
 CC gene or allele naturally present at a desired locus have at least two
 CC copies of the ORF, gene or allele, preferably in tandem arrangement.
 CC The modified bacteria are useful for producing chemical compounds.
 CC including L-amino acids, vitamins, nucleosides and nucleotides. The
 CC present sequence is encoded by a lysC gene, and represents an aspartate
 CC kinase carrying the T3111 mutation. Tandem duplications of the lysC
 CC gene are used to construct the bacteria of the invention.
 XX
 SQ Sequence 421 AA;
 Query Match 20.3%; Score 1227.5; DB 24; Length 421;
 Best Local Similarity 64.0%; Pred. No. 2.8e-85;
 Matches 251; Conservative 56; Mismatches 58; Indels 27; Gaps 5;
 QY 749 VQYGGSSVADAEIRRVAAERIVATKKQGNVWVVSAMGDTTDLDLAQVCPAPPPR 808
 Db 5 VQYGGSSLESARIRNVAERIVATKKAGNDVVVCSAMGDTTDLLELAANVPVPPAR 64
 QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPGRLQTA 868
 Db 65 EMDMLLTAGERISNALVAMAIESLGAESAQFTGSQAGVLTTERHGNARIVDVTGVRVREA 124

Qy	869	LEGRVVLVAGPQVSQDTKDVTTTLGRGSDTTAVAMAAALGADVCEIYTDVDGIFSADP	928
Db	125	LDEGKICIVAGPQVKNKTRDVTTLGRGSDTTAVAAAAALNADVCEIYSDVDGVYTADP	184
Qy	929	RIVNNAKLDVTAFEEMLNARCGAKVLMRCVEYARHNI PVHVRSSYSDBPGTVVVG	988
Db	185	RIVNNAKLEKLSFEEMLEAAVGSKILVRSVEYARAFNVPURVRSYSNDPGLIAG	244
Qy	989	IKDVPMEDPILTVGVAHDRSEAKVTIYGLPDIPGYAAKVFRAVA-----RRRQHR	1038
Db	245	MEDIPVEEAVLTGVATDKSEAKVTILGISDKPGEAAKVFRALADAEINIDMVLQNVSSVE	304
Qy	1039	HGAERLQGRGROD-RHLLHLP--QTSQPPPKWTRSETRSASTOLLYDDHIGKVS	1095
Db	305	DGTTDIIFTCPRSDGRRAMEILKLVQVG-----TNVLYDDQVGKVS	350
Qy	1096	GAGMRSHPGVTATFCEALAAVGVNIELISTSE	1127
Db	351	GAGKSHPGVTAEFMEALRDVNVNIELISTSE	382

Search completed: November 21, 2003, 16:03:16
 Job time : 73.9578 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 24.7156 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSVGNLRIA.....SAATRRPCTGRGRWACQ 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3810	62.9	741	3	US-09-001-984C-106
2	3810	62.9	741	4	US-09-396-347F-106
3	2570	42.4	731	4	US-09-252-991A-22524
4	2389	37.8	739	4	US-08-836-943-2
5	2225	36.7	729	4	US-09-328-352-7603
6	1545.5	25.5	421	4	US-08-311-731A-30
7	1223.5	20.2	421	1	US-08-532-828B-3
8	1223.5	20.2	421	1	US-08-700-359-9
9	1223.5	20.2	421	2	US-08-596-366-6
10	1223.5	20.2	421	3	US-08-967-104-6
11	1223.5	20.2	421	3	US-08-985-908-5
12	1223.5	20.2	421	3	US-08-852-730-14
13	1223.5	20.2	421	3	US-08-985-916-5
14	1219.5	20.1	421	1	US-08-532-828B-4
15	827	13.6	461	4	US-09-252-991A-23720
16	785	13.0	411	1	US-07-684-135A-2
17	770	12.7	156	4	US-09-056-556-227
18	770	12.7	156	4	US-09-072-598-222
19	746.5	12.3	438	4	US-09-328-352-6511
20	563	9.3	424	4	US-09-134-001C-5225
21	364.5	6.0	820	2	US-08-380-182-23
22	314.5	5.2	262	4	US-08-887-534A-23
23	314.5	5.2	262	4	US-09-527-431-23
24	313.5	5.2	863	2	US-08-380-182-19
25	310.5	5.1	449	1	US-08-256-136-2
26	310.5	5.1	449	2	US-08-950-737-2
27	310.5	5.1	449	2	US-08-973-461A-8

28 310.5 5.1 449 3 US-08-648-010-8 Sequence 8, Appli
29 310.5 5.1 863 2 US-08-380-182-20 Sequence 20, Appl
30 294.5 4.9 463 4 US-09-134-001C-4160 Sequence 4160, Ap
31 285.5 4.7 172 1 US-08-532-828B-5 Sequence 5, Appli
32 285.5 4.7 172 1 US-08-700-359-11 Sequence 11, Appl
33 285.5 4.7 172 3 US-08-985-908-7 Sequence 7, Appli
34 285.5 4.7 172 3 US-08-852-730-16 Sequence 16, Appl
35 285.5 4.7 172 3 US-08-985-916-7 Sequence 7, Appli
36 281.5 4.6 172 1 US-08-532-828B-6 Sequence 6, Appli
37 281.5 4.6 172 2 US-08-596-366-8 Sequence 8, Appli
38 281.5 4.6 172 2 US-08-967-104-8 Sequence 8, Appli
39 241 4.0 441 4 US-09-198-452A-1124 Sequence 1124, Ap
40 226 3.7 478 4 US-09-107-532A-6090 Sequence 6090, Ap
41 181.5 3.0 3892 4 US-09-328-352-5503 Sequence 5503, Ap
42 172.5 2.8 4545 2 US-08-804-227C-14 Sequence 14, Appl
43 172.5 2.8 4550 2 US-08-804-227C-8 Sequence 8, Appli
44 172.5 2.8 4550 2 US-08-804-198-2 Sequence 2, Appli
45 171 2.8 3798 3 US-09-335-409-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001,984C
; CURRENT FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; PRIOR FILING DATE: 1996-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-001-984C-106

Query Match 62.9%; Score 3810; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKXVADLTPQNALNARDELQ 68
Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKXVADLTPQNALNARDELQ 61
Qy 69 AQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDABITTTAGPQLVVPV 128
Db 62 AQIDKWHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDABITTTAGPQLVVPV 121
Qy 129 LNARFALNANARWGLYDALYGTDPVETPDGAEKGTYNKVRGDKVIARXFLDSDVP 188
Db 122 LNARFALNANARWGLYDALYGTDPVETPDGAEKGTYNKVRGDKVIARXFLDSDVP 181
Qy 189 LSSGSFGDAGTFTVDGQLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINHLHIEI 248
Db 182 LSSGSFGDAGTFTVDGQLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINHLHIEI 241
Qy 249 LIPDESQVGTDRAGVKDVILESAITTIMDFESVAADKVLGYRNWLGKNGDLAA 308
Db 242 LIPDESQVGTDRAGVKDVILESAITTIMDFESVAADKVLGYRNWLGKNGDLAA 301
Qy 309 AVDXDGFATFLRVLNDRNRYTAPGGQFTLPGRSLMFVRNVGHMTNDAI VDTDGSEVFEG 368
Db 302 AVDXDGFATFLRVLNDRNRYTAPGGQFTLPGRSLMFVRNVGHMTNDAI VDTDGSEVFEG 361

369 QY IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELSRVEDVLG 428
362 DB IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELSRVEDVLG 421
429 QY LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFTINTGFLDRTGDEIHTSMAGPMVRKG 488
422 DB LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFTINTGFLDRTGDEIHTSMAGPMVRKG 481
489 QY TMSQPMILAYEDHNVDAAGSAGRAQVKGKMTMTLMADMTETKIAQPRAGASTAW 548
482 DB TMSQPMILAYEDHNVDAAGSAGRAQVKGKMTMTLMADMTETKIAQPRAGASTAW 541
549 QY VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRBEVDNN 608
542 DB VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRBEVDNN 601
609 QY QCSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 668
602 DB QCSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 661
669 QY LERMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRRREF 728
662 DB LERMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRRREF 721
729 QY KARAAEKPAPESTRAGDDAAR 748
722 DB KARAAEKPAPESTRAGDDAAR 741

RESULT 2

US-09-396-347F-106
; Sequence 106, Application US/09396347F
; Patent No. 6506384
; GENERAL INFORMATION:
; APPLICANT: Laal, Suman
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Belisle, John T.
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: 32004-169276
; CURRENT APPLICATION NUMBER: US/09/396,347F
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 09/001,984
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-396-347F-106

Query Match 62.9%; Score 3810; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
69 AQIDKWHRRVIEPIIMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 128
62 AQIDKWHRRVIEPIIMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 121
129 LNARFALNANARWGSILYDALYGTVDVPEITDGAEGKPTNKVGRGDKVIAYAKFDDSDVP 188
122 LNARFALNANARWGSILYDALYGTVDVPEITDGAEGKPTNKVGRGDKVIAYAKFDDSDVP 181
189 LSGSGFGDATGFTVQDQQLVWALPKDSTGLANPGQFAGYTGAAEPTSVLLINHLGHEI 248
182 LSGSGFGDATGFTVQDQQLVWALPKDSTGLANPGQFAGYTGAAEPTSVLLINHLGHEI 241
249 LIDPESQVGTTRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA 308

242 DB LIDPESQVGTTRAGVKDVILSAITTIMDFEDSVAAVDAADKVLGYRNWGLNKGDLAA 301
309 QY AVDXDGTAFILVLRNDRNYTAPGGQFTLPGRSILMFVRNVGHLMTNDIAIVDTGSEVFE 368
302 DB AVDXDGTAFILVLRNDRNYTAPGGQFTLPGRSILMFVRNVGHLMTNDIAIVDTGSEVFE 361
369 QY IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELSRVEDVLG 428
362 DB IMDALFTGLIAIRGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELSRVEDVLG 421
429 QY LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFTINTGFLDRTGDEIHTSMAGPMVRKG 488
422 DB LPQNTWKIGIMDBERTTNLAKACIKAAADRVVFTINTGFLDRTGDEIHTSMAGPMVRKG 481
489 QY TMSQPMILAYEDHNVDAAGSAGRAQVKGKMTMTLMADMTETKIAQPRAGASTAW 548
482 DB TMSQPMILAYEDHNVDAAGSAGRAQVKGKMTMTLMADMTETKIAQPRAGASTAW 541
549 QY VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRBEVDNN 608
542 DB VPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRBEVDNN 601
609 QY QCSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 668
602 DB QCSILGYVVRWVDOGVGCKSKVPDIHVALMEDRATLRISSOLLANWLHGVITTSADVRAS 661
669 QY LERMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRRREF 728
662 DB LERMAPLVDRQAGDVAYRPMAPNFDDSIATFLAAQELILSGAQPNNGYTEPILHRRRREF 721
729 QY KARAAEKPAPESTRAGDDAAR 748
722 DB KARAAEKPAPESTRAGDDAAR 741

RESULT 3

US-09-252-991A-22524
; Sequence 22524, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22524
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22524

Query Match 42.4%; Score 2570; DB 4; Length 731;
Best Local Similarity 69.1%; Pred. No. 9.6e-222;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;

9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
8 TERVQVGLQVAKVLPDFVNNALPGTGVASDFTWTAAGAVINDLAPKNKALLAKXDELQ 67
69 AQIDKWHRRVIEPIIMDAYRQPLTEIGYLLPEPDDFTITTSVDAEITTTAGPQLVWPV 128
68 AKIDGWHQARAGQAHAAYAKFLEEIGYLLPEADFOAGTQNVDDDEIARMAGPQLVWPV 127
129 LNARFALNANARWGSILYDALYGTVDVPEITDGAEGKPTNKVGRGDKVIAYAKFDDSDVP 188
128 MNARFALNANARWGSILYDALYGTVDVISEGGGKCKGKYNKVRGDKVIAFAFLDEAAP 187

189 LSSGSGDATGTTVODGQVVALPKS--TGLANPCQFAGYTCAAESPTSVLLINHLHIE 247
189 LSSGSHVATSVKNGALVWALKNGSETGLKNAQFLAFQGDAAKCAVLLKINGLHFE 247
248 ILIDPESQVTTDRAGVKOVILESAITTIMDPESVAAVDAADKVLGVRNWLKNGKGLA 307
248 IQIDPSSVPGQDAGVNDVLEAALITTIMDCESVAAVDAADKVLGVRNWLKNGKGLA 307
308 AAVDGDGTAFRLVRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNTDAIVDTDGSVEFE 367
308 BEVSKGSGTFTTVMPPDVAVYTRADGSELT.LHGRSLLFVRNVGHLMNTDAILLKQNEVPE 367
368 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSTYIVKPKMHGPAEVAFTCELFPSRVED 425
368 GIQDGLFTSLAIH-----DLNGNTRSRSRTSGSYIVKPKMHGPEEAAFTNELFGRVED 422
426 VLGLPONTMKGIMDEEBERTTNLKACIKAAADRVVFINTGFLORTGDEIHTSWEAGPMV 485
423 VLGLPRNTLKVIMDEEBERTTNLKACIKAAKDRVVFINTGFLORTGDEIHTSWEAGAVV 482
486 RKGTMKSQFWILAYEDHNVDAGLAAGFSRAQVKGGMWMTMELMADVETKXIAQPRAGAS 545
483 RKGAMKSEKIGAVENNVVGLATGLQKQAIQKGMWAMPDLMAAMLEQKIGHPLAGAN 542
546 TAWPSPPTAATHALHYHOVAVAAVOOGLAGKRRATIEQLLTIPLAKELAWAPDEIRREV 605
543 TAWPSPPTAATHALHYHKVDVFAQAEIAKRTFASVDDIITIPLAPNTNWTABEIKNEV 602
606 DNNCOSILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGVITGADV 665
603 DNNAGILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGVISQEQV 662
666 RASLERMAPLVDQRNAGVAVRMAPNPDSDIAFLAQELILSGAQQPNGYTEPILHRRR 725
663 VESLKRMAVVVDQRNASDPSPRMAPNPDSDIAFLAQELILSGAQQPNGYTEPILHRRR 722
726 REFRAK 731
723 REFRAK 728

RESULT 4

US-08-836-943-2
; Sequence 2, Application US/08836943
; Patent No. 5965391
; GENERAL INFORMATION:
; APPLICANT: Reinscheid, Dieter
; APPLICANT: Eikmanns, Bernhard
; APPLICANT: Sahm, Hermann
; TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN
; TITLE OF INVENTION: CORYNEFORM BACTERIA
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Firm of Karl F. Ross, PC
; STREET: 5676 Riverdale Ave.
; CITY: Bronx
; STATE: New York
; COUNTRY: USA
; ZIP: 10471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,943
; FILING DATE: 08-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Jonathan
; REGISTRATION NUMBER: 26,963
; REFERENCE/DOCKET NUMBER: 20357
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (718) 884-6600
; TELEFAX: 718/601-1099
; TELEX: 620428
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-836-943-2

Query Match 37.8%; Score 2289; DB 2; Length 739;
Best Local Similarity 61.3%; Pred. No. 1.7e-196;
Matches 444; Conservative 90; Mismatches 184; Indels 6; Gaps 4;

QY 9 TDSVSVGNLRIRAVLYDFVNNALPGCTDIDPDSFWAGVDKVVADLTPOQALLNADEIQ 68
DB 20 TSVVDAGGQVAKVLYDFTEAVLPVGVDAKFSGFAIARDLTPFRRELLARDEIQ 79
QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVVPV 128
DB 80 MLIDDYHRNN-SCGIDQEAIVEDFLKEIGYLVPEEPAEAEIRTNQVDTESISSTAGPQLVVP 138
QY 129 LNARFALNAANARWGLSYDALYGTVDVLPETDGAEKQPTNKVGRDXVIAYAKFLDSDVP 188
DB 139 LNARFALNAANARWGLSYDALYGTNAIPETDGAEKQPTNKVGRDXVIAYAKFLDSDVP 198
QY 189 LSSGSGDATGFTVODGQVVALPKDSTGLANPCQFAGYTCAAESPTSVLLINHLHIEI 248
DB 199 LDCASHADYVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEALLLETNGHLIEL 258
QY 249 LIDPESQVTTDRAGVKOVILESAITTIMDPESVAAVDAADKVLGVRNWLKNGKGLAA 308
DB 259 QIDPVPHPGKADTGHKDVLESATITIMDPESVAAVDAADKVLGVRNWLKNGKGLAA 318
QY 309 AVDKDGTAFRLVRNDRNYTAPGGQFTLPGRSLMFVRNVGHLMNTDAIVDTDGSVEFE 368
DB 319 EMSKNGRIFTRELNKDRVYIGRNGTELVHGRSLLFVRNVGHLMNQPSIL-IDGEEIFEG 377
QY 369 IMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELFPSRVEDVLG 428
DB 378 IMDALFTVCAIPGIAPQN---KMRNSRKGSYIVKPKMHGPAEVAFTCELFPSRVEDLLD 434
QY 429 LPONTMKGIMDEEBERTTNLKACIKAAADRVVFINTGFLDRGTDEIHTSMENGPVRKG 488
DB 435 LPRHTLKGVNDEEBERTSVNLDASINMEVADRIAFINTGFLDRGTDEIHTSMENGPVRKA 494
QY 489 TMSQFWILAYEDHNVDAGLAAGFSRAQVKGGMWMTMELMADVETKXIAQPRAGASTAW 548
DB 495 DMQTAFWKQAYENNVNVDAGIQRGLPKCAQIKGGMWMTMELMADVETKXIAQPRAGASTAW 554
QY 549 VPSPTAATHALHYHOVAVAAVOOGL-AGKRRATIEQLLTIPLAKELAWAPDEIRREV 607
DB 555 VPSPTGATLHATHYHLVDYFKVQDELRAAGRRDLSLENILTIPTAPNTNSEEKKEMDN 614
QY 608 NCOSILGYVVRVWDQVGGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGVITSDVRA 667
DB 615 NCOSILGYVVRVWEHGVGSKVPDIHDVALMEDRATLRSSOLLANLWLRHGVITSDVRA 674
QY 668 SLERMAPLVDQRNAGVAVRMAPNPDSDIAFLAQELILSGAQQPNGYTEPILHRRR 727
DB 675 SLERMAVVVDQRNASDPSPRMAPNPDSDIAFLAQELILSGAQQPNGYTEPILHRRR 734
QY 728 FKAR 731
DB 735 FKAR 738

RESULT 5

US-09-328-352-7603
; Sequence 7603, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7603
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7603

Query Match 36.7%; Score 2225; DB 4; Length 729;
Best Local Similarity 59.8%; Pred. No. 9.5e-191;
Matches 432; Conservative 97; Mismatches 185; Indels 8; Gaps 4;

QY 9 TDRVSGNRIARVLYDFVNNALPGTDIDPDSFWAGYDKVADVLTPOQALLNARDLEQ 68
DB 10 TARIQKLAIAKELYDFIENEALPGSLDSEYTKVQVVDLSPKNALLAKRDELQ 69

QY 69 AQIDKWHRRVIRPIDMDAYRQPLTIGVLLPEPDDFTTTSQVDAEITTTAGPOLVVPV 128
DB 70 AKIDEWHRRNKF--LGAYKAFLEITIGVLLPEVEDFQITTNVDEEIALLAGPOLVVPV 126

QY 129 LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAIYARKFLLDDSV 188
DB 127 RNARYCLNANARWGLYDALYGFVISEBGAEGKGYNPRVGAKEVFAEFKNFLNEIFP 186

QY 189 LSSGSGDATCFVQDQGVVLPD-KSTGLANPGOFAGYTGAAESPTSVLLINGLHIE 247
DB 187 LAQGSHADATKYAEQKLVLT-KDTGTTGLAHEAGVGFNGEENAFSEVLLSGLHVI 246

QY 248 ILIDPESQVTTDRAGVKDVILESALTITMDPESVAAVDAADKVLGYRNWGLNKGDI 307
DB 247 IEIDANSPIGQTLAGVKOLTLEAVTTIQDLEDSVAAVDAEBKVEGYENWGLNKGTLQ 306

QY 308 AAVDKGTAPLRVLRDRNVTAPGGQFTLPQSLMFVRNVGHMTNDIAVITDQSEVPE 367
DB 307 ESIEKNGKTIIVRALNDRKREIKNLIGGTTKLHGRSLMLLRNVGHMTNPAIL-VDGEIIFE 365

QY 368 GIMDALFTGLIATHGLKASDVNGPLNSRTGSIYIVKPKVHGPVAVFTCELPFSVEDVL 427
DB 366 GIMDALVTPLLSLTADIRSENVK--NSRKGSMYIVKPKVHGPVAVFAVELFERAQL 422

QY 428 GLPONTWIKIGIMDEERTTNLAKACIAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
DB 423 GLPAKSLKIGIMDEERTSNLKNCAAAAKDRITFINTGFMDBRTGDEIHTSMEAPVVRK 482

QY 488 GTWKSOPWILAYEDHNVDAAGAFSGRAQVKGKMTMTLMADNVETKIAOPRAGASTA 547
DB 483 EAVKTKWIAAYENRNVAILGKCGLOGKQAGIKGKMPKPSMKMDLMTAAAPNAGASCA 542

QY 548 WPSPTAATLHALHYHQVDVAAVQOGLAGKRATIEQLLTIPLAKELAWAPDEIEEVDN 607
DB 543 WPSPTGAVLHAGHYHQVNVKARQDQKAEEMLSLDLLTPPATDTNWSAEIENNELN 602

QY 608 NCQSILGVYVWVDQGVGSKVPDIHDVALMEDRATIRISSOLLANLWLRHGVITADVPA 667
DB 603 NCQGIILGVYVWVDLGVGSKVPDIINNVGLMEDRATIRISSQHVANLWLRHGVITREQVEE 662

QY 668 SLERMAFLVDRQAGVAVRPMAPNPFDSIAFLAAQBELISLAQOQNGVTEPILHRRRRE 727
DB 663 VLKEMAKIVDEQNANDPLYKPMANFETNIAFOAASDLIFKCEQPSGYTERELLHAARLK 722

QY 728 FK 729
DB 723 LK 724

RESULT 6
US-08-311-731A-30
; Sequence 30, Application US/08311731A

; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LEPRAE
US-08-311-731A-30

Query Match 25.5%; Score 1545.5; DB 4; Length 421;
Best Local Similarity 80.4%; Pred. No. 4.2e-130;
Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 749 VQYGGSSVADAERIRRAERIVATKQGNVWVWSAMGDTTDDLLDLAQVCPAPP 808
DB 5 VQYGGSSVADADRIRRAERIVQTKQGNDIVVWSAMGDTTDDLLDLAQVCPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGHNAKIIDVTPLQTA 868
DB 65 ELDMLLTAGERISNALVAMAIESFGAARSFTGSOAGVITTTGHNAKIIDVTPLQTA 124

QY 869 LEEGRVVLVAGFQGVSDTQVTTGLSGGSDTTAVAVAAALGADVCEIYTDVDFISADP 928
DB 125 LDEGRVVLVAGFQGVSDTQVTTGLSGGSDTTAVAVAAALGADVCEIYTDVDFISADP 184

QY 929 RIVNARAKLDTVFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSISDRGTGVVGS 988
DB 185 RVVFNARAKLDTVFEEMLEMAACGAKVLMRCVEYARRHNIPIVHVRSSISDRGTGVVGS 244

QY 989 IKDVPMDPILTVGAHDSERAKVTIVGLPDI PGVAAKVFRAVARRRQHRHGAEE----- 1043
DB 245 IKDVPMDPILTVGAHDSERAKVTIVGLPDI PGVAAKVFRAVARRRQHRHGAEE----- 295

QY 1044 RLQGRGQDRHHLHL-----LPQSGPPPKWKTSETRSASTQLLYDDHIGKVLIGAGMR 1100
DB 296 VLQNVKVEDGKTDITFTCSRDSGPVAVAKGLSLRDEIGFTQLLYDDHIGKVLIGAGMR 355

QY 1101 SHPGVTATFEALAAVGVNIELISTSEDSORSCCAAT 1137
DB 356 SHPGVTATFEALAAVGVNIELISTSEDSORSCCAAT 392

RESULT 7
US-08-532-828B-3
; Sequence 3, Application US/08532828B
; Patent No. 5688671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,828B
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-101450
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-764-0 PCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Corynebacterium glutamicum
; STRAIN: ATCC13869
; US-08-532-828B-3

Query Match 20.2%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLDDLAQQVCPAPPPR 808
Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELLEAAVNPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMATESLGAHARSFTGSGAQTGTHGNAKIIDVTPGRLQTA 868
Db 65 EMDMLLTAGERISNALVAMATESLGAESAQSFSGAQLTTERHGNARIVDVTGVRVREA 124
QY 869 LEEGRVVLVAGFQGVSDTKVTTILGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928
Db 125 LDESKI CIVAGFQGVKNETRVTTILGRGSDTTAVAAALNADVCEIYSDVGVYADP 184
QY 929 RIVNARKLDTVTPEEMLEMAACAKVLMFLRCVEYARRHNPVHRVRSYSYDRPGTVVVG 988
Db 185 RIVENAQKLEKLSPEEMLELAAGSKILVLRSEYARAFNVFLRVRSYSYNDPGLIAGS 244

QY 989 IKDVPMPEDPILTCVAHDRSEAKVTIVGLPDPGAAKVERAVA-----RRRQHR 1038
Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLSDKPEAAKVFALADAEINIDMWLVQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPPWKNWTRSETRSASTQLLYDDHIGKVSLLI 1095
Db 305 DGTDTITFTCPRADGRAMEILKQVQG-----TNVLYDDQGVKVSIV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
Db 351 GAGMKSHPGVTAFMEALRDNVNVNIELISTSE 382

RESULT 8
US-08-700-359-9
; Sequence 9, Application US/08700359
; Patent No. 5766925
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, MASAKAZU
; APPLICANT: USUDA, YOSHIHIRO
; APPLICANT: SUZUKI, TOMOKO
; APPLICANT: TANAKA, AKIKO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS TEXT EDITOR
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,359
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-35019
; FILING DATE: 04-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-819-0 PCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-700-359-9

Query Match 20.2%; Score 1223.5; DB 1; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLDDLAQQVCPAPPPR 808
Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELLEAAVNPVPPAR 64
QY 809 ELDMLLTAGERISNALVAMATESLGAHARSFTGSGAQTGTHGNAKIIDVTPGRLQTA 868
Db 65 EMDMLLTAGERISNALVAMATESLGAESAQSFSGAQLTTERHGNARIVDVTGVRVREA 124
QY 869 LEEGRVVLVAGFQGVSDTKVTTILGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928

Db 125 LDEGKICIVAGFGQVGNKETRDVTTLGRGSDTTAVALAAALNADVCEIYSDVDGVYVTDAP 184
QY 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 988
Db 185 RIVPNAQKLEKLSFEEMLEAAVGVSKILVRSVEYARAFNPLRVRSYSNDPGTLLIAGS 244
QY 989 IKDVPNEDPILTCVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPCEAAKVFRALADAEINIDMWLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPPKWNTSRSETSASTQLLYDDHIGKVSLLI 1095
Db 305 DGTDTITFCPRADGRAMEILKKLVQVQ-----NW-----TNVLYDDQVGVKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 9

US-08-596-366-6
; Sequence 6, Application US/08596366
; Patent No. 5876983
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/596,366
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-784-0 PCT
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-596-366-6

Query Match 20.2%; Score 1223.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 3 6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;
QY 749 VQYGGSSVADAEIRIRVAERIVATKKQGNVNVVVSAMGDTTDDLLDLAQVCPAPPPR 808
Db 5 VQYGGSSLESASRIENVAERIIVATKKAGNDVVVVCAMGDTTDELELEAAVNPVPPAR 64
QY 809 ELMWLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTGSLQTA 868
Db 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNARIVDVTGVRRA 124
QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSDTTAVAMAALGADVCEIYTDVDFISADP 928
Db 125 LDEGKICIVAGFGQVGNKETRDVTTLGRGSDTTAVALAAALNADVCEIYSDVDGVYVTDAP 184
QY 929 RIVRNARKLDTVTTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVGVS 988
Db 185 RIVPNAQKLEKLSFEEMLEAAVGVSKILVRSVEYARAFNPLRVRSYSNDPGTLLIAGS 244
QY 989 IKDVPNEDPILTCVAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
Db 245 MEDIPVEEAVLTGVATDKSEAKVTIVGLISDKPCEAAKVFRALADAEINIDMWLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPPKWNTSRSETSASTQLLYDDHIGKVSLLI 1095
Db 305 DGTDTITFCPRADGRAMEILKKLVQVQ-----NW-----TNVLYDDQVGVKVSLLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 10
US-08-967-104-6
; Sequence 6, Application US/08967104
; Patent No. 5919694
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: IZUI, Katsura
; TITLE OF INVENTION: MUTANT PHOSPHENOLPYRUVATE CARBOXYLASE,
; TITLE OF INVENTION: ITS GENE, AND PRODUCTION METHOD OF AMINO ACID
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER, & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/967,104
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,366
; FILING DATE: 29-APR-1996
; APPLICATION NUMBER: JP 5-209775
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-209776
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-153876
; FILING DATE: 05-JUL-1994

ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-784-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-104-6

Query Match 20.2%; Score 1233.5; DB 2; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLDAQQVCPAPPPR 808
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIDVTTCRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNARIIVDTTCRVREA 124

QY 869 LEEGRVVLVAGFGQVQSDTKVTLGRGSDTTAVAMAAALGADVCEIYTDVDFGIFGADP 928
DB 125 LDEGKICIVAGFGQVQNETRDVTLGRGSDTTAVALLAALNADVCEIYSDVDGVYTDAP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDEGTVVVGS 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLVRSSYSNDPDTLIAGS 244

QY 989 IKDVPMDPILTCVADHRSSEAKVTIVGLPDI PGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTIVGLISDKEGAQVFRALADAENIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHLLHLP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 1095
DB 305 DGTDTITTCPRADGRAMEILKQLVQV-----NW-----TNVLYDDQVGKVS LV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 11
US-08-968-5
Sequence 5, Application US/08985908
Patent No. 6004773
GENERAL INFORMATION:
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIMURA, AND TSUYOSHI NA
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,908
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-325659
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-908-5

Query Match 20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLDAQQVCPAPPPR 808
DB 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVCSAMGDTTDELELAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNAKIIDVTTCRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAEISLGAHARSFTGSAQVITTTGTHGNARIIVDTTCRVREA 124

QY 869 LEEGRVVLVAGFGQVQSDTKVTLGRGSDTTAVAMAAALGADVCEIYTDVDFGIFGADP 928
DB 125 LDEGKICIVAGFGQVQNETRDVTLGRGSDTTAVALLAALNADVCEIYSDVDGVYTDAP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDEGTVVVGS 988
DB 185 RIVPNAQKLEKLSFEEMLELAAGVKILVLRSEYARAFNPLVRSSYSNDPDTLIAGS 244

QY 989 IKDVPMDPILTCVADHRSSEAKVTIVGLPDI PGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGATDKSEAKVTIVGLISDKEGAQVFRALADAENIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHLLHLP--QTSPPPPKWNTRSETRSASTQLLYDDHIGKVS LI 1095
DB 305 DGTDTITTCPRADGRAMEILKQLVQV-----NW-----TNVLYDDQVGKVS LV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 12
US-08-852-730-14
Sequence 14, Application US/08852730
Patent No. 6090597
GENERAL INFORMATION:
APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EIICHI NAKANO,
APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIMURA, AND TSUYOSHI
APPLICANT: NAKAMATSU
TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
ZIP: 22026
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,730
FILING DATE: 05-07-1997

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: JP 8-142812
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-730-14

Query Match 20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADARRIRVAERIVATKKQGNVVVVSAMGDTTDDLDDLAQVCPAPP 808
DB 5 VQYGGSSLESARIRVAERIVATKKAGNDVVVCSAMGDTTDELELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVMAIESLGAESAQFTGSOAGVLTTERHGNARIVDVTPGRVREA 124
QY 869 LEGRVVLVAGFGQVSQDVTTLGRGSDTTAVAMAAALGADVCEIITVDVGIFSDAP 928
DB 125 LDEGKICIVAGFGQVKNKTRDVTTLGRGSDTTAVAAAAALNADVCEIIVSDVDGVYADP 184
QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVG 988
DB 185 RIVRNAQKLEKISFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSNDPGTLLIAGS 244
QY 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGVATDKSEAKVTVLGIDSKPGEAAKVFALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSEASQTLVYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPADGRAMEILKLVQVG-----TNVLYDDQVGVKSLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAFMEALRDVNVNIELISTSE 382

RESULT 13
US-08-985-916-5
; Sequence 5, Application US/08985916
; Patent No. 6221636
; GENERAL INFORMATION:
; APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; ZIP: 22152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,916
; FILING DATE: 05-DEC-1997

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: JP 8-325658
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 421 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-916-5

Query Match 20.2%; Score 1223.5; DB 3; Length 421;
Best Local Similarity 64.0%; Pred. No. 3.6e-101;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADARRIRVAERIVATKKQGNVVVVSAMGDTTDDLDDLAQVCPAPP 808
DB 5 VQYGGSSLESARIRVAERIVATKKAGNDVVVCSAMGDTTDELELAANVPVPPAR 64
QY 809 ELDMLLTAGERISNALVMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLLTAGERISNALVMAIESLGAESAQFTGSOAGVLTTERHGNARIVDVTPGRVREA 124
QY 869 LEGRVVLVAGFGQVSQDVTTLGRGSDTTAVAMAAALGADVCEIITVDVGIFSDAP 928
DB 125 LDEGKICIVAGFGQVKNKTRDVTTLGRGSDTTAVAAAAALNADVCEIIVSDVDGVYADP 184
QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNIPVHVRSSYSDRPGTVVVG 988
DB 185 RIVRNAQKLEKISFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSNDPGTLLIAGS 244
QY 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRQHR 1038
DB 245 MEDIPVEEAVLTGVATDKSEAKVTVLGIDSKPGEAAKVFALADAEINIDMVLQNVSSVE 304
QY 1039 HGAERLQGRQD-RHHLHLLP--QTSGPPPKWNTSRSEASQTLVYDDHIGKVSLLI 1095
DB 305 DGTDTITFTCPADGRAMEILKLVQVG-----TNVLYDDQVGVKSLV 350
QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 351 GAGMKSHPGVTAFMEALRDVNVNIELISTSE 382

RESULT 14
US-08-532-828B-4
; Sequence 4, Application US/08532828B
; Patent No. 5688671
; GENERAL INFORMATION:
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: OGAWA, Yuri
; APPLICANT: SUZUKI, Tomoko
; APPLICANT: TANAKA, Akiko
; APPLICANT: NATSUI, Hiroschi
; TITLE OF INVENTION: MUTANT ASPARTOKINASE GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: MS-DOS TEXT EDITOR
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/532,828B
 FILING DATE: 27-OCT-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 5-101450
 FILING DATE: 27-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: NORMAN F. OBLON
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 10-764-0 PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 421 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Corynebacterium glutamicum
 STRAIN: AJ3463
 US-08-532-828B-4

Query Match 20.1%; Score 1219.5; DB 1; Length 421;
 Best Local Similarity 63.8%; Pred. No. 8.3e-101;
 Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;
 QY 749 VQYGGSSVADARIRRAERIVATKKGNDVVVVVVSANGDTTDDLLDLAQOVCAPP 808
 Db 5 VQYGGSSLESARIRNVAERIVATKAGNDVVVVVCSANGDTTDELLEAAVNPVPPAR 64
 QY 809 ELDMLLTAGERISNALVAVATESLCAHARSPTGSGAGVITGTGHNAKIIDVTGRLQTA 868
 Db 65 ENDMLLTAGERISNALVAVATESLGAESAQSTFGSGAGVITGTGHNARIVDVTGVRREA 124
 QY 869 LEEGRVILVAGQGVQSDTKVTTTLGRGSGDTTAVMAAALGADVCEIYTDVDDGIFSDAP 928
 Db 125 LDEGKICIVAGQGVQSKETRDVTTTLGRGSGDTTAVALLAALNADVCEIYSDVGVYADP 184
 QY 929 RIVRNARKLDTVTEEMLEMAACGAKVLMRCVEYARHNIPVHVRSSYSDRPGTVVVG 988
 Db 185 RIVPNAQKLEKUSFEEMLELAAGSKILVLRSEYARAFNPLVRSSYSDNDPCTLIAGS 244
 QY 989 IKDVPMDPILTVGAHDSRAKVTIVGLPDIPGVAAKVFRAVA-----RRRQHR 1038
 Db 245 MEDIPVEEAVLTGVATDKSEAKVTVLGSDKPGETAKVFRALADAEINIDMVLQNVSSVE 304
 QY 1039 HGAERLQGRQD-RHLLHLP--QTSGLPPKWNTRSETRSASTQLLYDDHIGKVSLLI 1095
 Db 305 DGTDTITFTCPRADGRRAEILKKLVQVG-----NW-----TNVLYDDQGVKVSIV 350
 QY 1096 GAGMSEHPGVATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGMKSHEFGVTAEFNEALRDVNVNIELISTSE 382

RESULT 15

US-09-252-991A-29720
 Sequence 29720, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 29720
 LENGTH: 461
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-29720

Query Match 13.6%; Score 827; DB 4; Length 461;
 Best Local Similarity 38.6%; Pred. No. 1.8e-65;
 Matches 187; Conservative 86; Mismatches 130; Indels 82; Gaps 7;
 QY 663 ADVRASLERMAPLVDRQCNAGDVYRMAFNFDSDIAFLAAQELLISGAQOQNGYTPILH 722
 Db 6 APERCAGRPAPLDARP-----AVRPASP-----GSENMWQVQKALH 43
 QY 723 RRRREFKARAAEXPAFSDRAGDDAARVQKYGGSSVADAERIRRAERIVATKKGNDVVV 782
 Db 44 GIRR-----FEMALIVQKFGTSGVTVERIEQVAEKVKFREAGDDVVV 87
 QY 783 VVSAMGDTTDDLLDLAQOVCAPPRELDMLLTAGERISNALVAVATESLCAHARSFTGS 842
 Db 88 VVSAMGETNRLTGLANQIMEQVPRELDVVVSTGEQVTIALLSMALIKEGVPAVSYTGN 147
 QY 843 QAGVITGTGHNAKIIDVTGRLQTALEGRVVLVAGFQGVSDTKDVTTLGRGSGDTTA 902
 Db 148 QVAILTDSATKARILHIDDTHTIRADLKAGRVVVVAGFQGV-DGNGNITTLGRGSGDTTG 206
 QY 903 VMAAALGADVCEIYTDVDDGIFSDAPRIVRNARKLDTVTFEEMLEMAACGAKVLMRCVE 962
 Db 207 VALAAALKADECIYTDVDDGVYTTDPRVVPQARRLDKITFEEMLEMASLGSVKVQIRAVE 266
 QY 963 YARRHNIPVHVRSSYSDRPGTVVVGSIKDVPMEDPILTVGAHDSRAKVTIVGLPDIPGY 1022
 Db 267 FAKYINVPLRVLSFQEGPGTLLTIDDESEMEQPIISGTAFNREDAKLTIRGVDPDPGV 326
 QY 1023 AAKVFRAVARRRQ-----HRHGAERLQGRQDRHLLHLPQT 1062
 Db 327 AFKILGPISAANVEVDIMVQNVAHDNTDFTFTVHRNDVNLAL-----ILKQT 375
 QY 1063 SGPPPKWNTRSETRSASTQLLYDDHIGKVSLLIGAGRSHPGVYATFCEALAAVGVNIEL 1122
 Db 376 A-----ANIGAREAIGDTNIAKSVIVGVGNRSHAGVASRNFALAKESINIQM 423
 QY 1123 ISTSE 1127
 Db 424 ISTSE 428

Search completed: November 21, 2003, 16:11:41
 Job time : 26.7156 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 45.7107 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSVGNLRIA.....SAATRRPCTRGDRGWACQ 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 18259486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3868	63.8	748	11	US-09-272-975-2
2	3868	63.8	753	11	US-09-272-975-58
3	2289	37.8	739	10	US-09-738-626-6988
4	1615.5	26.7	421	9	US-09-793-366-163
5	1224.5	20.2	421	10	US-09-738-626-3781
6	1224.5	20.2	421	11	US-09-746-660A-24
7	1223.5	20.2	421	15	US-10-226-136-5
8	1217.5	20.1	421	15	US-10-067-974-2
9	1211.5	20.0	421	15	US-10-067-974-16
10	1144.5	18.9	430	15	US-10-156-761-12093
11	770	12.7	156	12	US-10-084-843-227
12	770	12.7	156	12	US-10-193-002-222
13	673	11.1	394	12	US-09-882-227-290
14	377	6.2	564	10	US-09-890-813-17
15	365	6.0	555	10	US-09-890-813-8

16	359	5.9	560	10	US-09-890-813-6	Sequence 6, Appli
17	314.5	5.2	262	12	US-10-301-997-23	Sequence 23, Appl
18	310.5	5.1	449	11	US-09-989-339-23	Sequence 23, Appl
19	306.5	5.1	439	10	US-09-890-813-16	Sequence 16, Appl
20	285.5	4.7	172	15	US-10-226-136-7	Sequence 7, Appli
21	248.5	4.1	281	10	US-09-890-813-12	Sequence 12, Appl
22	241	4.0	440	15	US-10-022-832-34	Sequence 34, Appl
23	197	3.3	4999	10	US-09-976-059-14	Sequence 14, Appl
24	184	3.0	530	15	US-10-156-761-9582	Sequence 9582, Ap
25	179	3.0	304	15	US-10-156-761-14293	Sequence 14293, A
26	173	2.9	5245	12	US-10-329-079-11	Sequence 11, Appl
27	171	2.8	3798	14	US-10-014-717-6	Sequence 6, Appli
28	168	2.8	3970	15	US-10-156-761-10429	Sequence 10429, A
29	166	2.7	3564	15	US-10-156-761-7964	Sequence 7964, Ap
30	161	2.7	6395	12	US-09-940-316B-72	Sequence 72, Appl
31	160	2.6	5245	12	US-10-329-079-45	Sequence 45, Appl
32	151.5	2.5	7746	15	US-10-156-761-7965	Sequence 7965, Ap
33	150	2.5	6146	15	US-10-156-761-10436	Sequence 10436, A
34	149.5	2.5	11877	10	US-09-861-289-6	Sequence 6, Appli
35	149.5	2.5	11877	11	US-09-860-846-6	Sequence 6, Appli
36	149.5	2.5	11877	11	US-09-836-821-6	Sequence 6, Appli
37	149.5	2.5	11877	12	US-10-271-889-49	Sequence 49, Appl
38	149.5	2.5	12199	11	US-09-988-384B-6	Sequence 6, Appli
39	148.5	2.5	4840	15	US-10-156-761-10435	Sequence 10435, A
40	146	2.4	3352	15	US-10-156-761-7961	Sequence 7961, Ap
41	146	2.4	7257	14	US-10-014-717-5	Sequence 5, Appli
42	145.5	2.4	541	15	US-10-156-761-9538	Sequence 9538, Ap
43	145	2.4	1624	16	US-10-080-170-539	Sequence 539, App
44	145	2.4	8360	12	US-10-132-134-34	Sequence 34, Appl
45	144.5	2.4	2472	9	US-09-815-242-5064	Sequence 5064, Ap

ALIGNMENTS

RESULT 1
US-09-272-975-2
; Sequence 2, Application US/09272975
; Publication No. US2003002774A1
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-2

Query Match	63.8%	Score	3868	DB	11	Length	748
Best Local Similarity	100.0%	Pred. No.	0				
Matches	748	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MQHHHHHTDRVSVGNLRIARVLYDFVNEALPGTIDPDSFWAGYKVVADLTPOQAL	60				
Db	1	MQHHHHHTDRVSVGNLRIARVLYDFVNEALPGTIDPDSFWAGYKVVADLTPOQAL	60				
Qy	61	LNARDELQIQDKWHRRTIEPIDMDAYQFTEIGYLLPEPDDFTTTSGVDAEITTTA	120				
Db	61	LNARDELQIQDKWHRRTIEPIDMDAYQFTEIGYLLPEPDDFTTTSGVDAEITTTA	120				
Qy	121	GPOLVVPVNLNARPALNANRGSGLYDLYGTVDVIPETDGAEGKPTYNKVRGDKVIAYAR	180				
Db	121	GPOLVVPVNLNARPALNANRGSGLYDLYGTVDVIPETDGAEGKPTYNKVRGDKVIAYAR	180				
Qy	181	KFLDDSVPLSSGSGFGATGTFTVDGOLVVALPDKSTGLANPGOFAGYTGAEPSPTSVLLI	240				

```
Db 181 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSLLI 240
Qy 241 NHGLHIEILIDPSQVGTTRAGVKDVIIESAITTIMDFEDSVAADAAKVLGYRNWLG 300
Db 241 NHGLHIEILIDPSQVGTTRAGVKDVIIESAITTIMDFEDSVAADAAKVLGYRNWLG 300
Qy 301 LNKGDIAAAVDKGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIIVDT 360
Db 301 LNKGDIAAAVDKGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIIVDT 360
Qy 361 DGSEVPEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Db 361 DGSEVPEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Qy 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKCACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Db 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKCACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Qy 481 AGPMVRKGTMTKSPWILAYEDHNVDAAGSRAQVKGKMTMTLMADVMVETKIAQP 540
Db 481 AGPMVRKGTMTKSPWILAYEDHNVDAAGSRAQVKGKMTMTLMADVMVETKIAQP 540
Qy 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Db 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Qy 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSELLANLWLRHGVI 660
Db 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSELLANLWLRHGVI 660
Qy 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPI 720
Db 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPI 720
Qy 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
Db 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
```

RESULT 2

```
US-09-272-975-58
; Sequence 58, Application US/09272975
; Publication No. US2003002774A1
```

GENERAL INFORMATION:

```
; APPLICANT: Hendrickson, Ronald C.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: TUBERCULOSIS ANTIGENS AND METHODS
; FILE REFERENCE: 210121.474
; CURRENT APPLICATION NUMBER: US/09/272,975
; CURRENT FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-272-975-58
```

```
Query Match 63.8%; Score 3868; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 748; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQHHHHHTDRVSGNLRARVLYDFVNEALPGTIDIDPSFWAGVKVADLTQNAL 60
Db 6 MQHHHHHTDRVSGNLRARVLYDFVNEALPGTIDIDPSFWAGVKVADLTQNAL 65
Qy 61 LNARDELQAIQDKWHRRRIEPTDMDAYRQFLTEIGYLLPEPDDFTITTSQVDASITTTA 120
Db 66 LNARDELQAIQDKWHRRRIEPTDMDAYRQFLTEIGYLLPEPDDFTITTSQVDASITTTA 125
```

```
Qy 121 GPQLVPLVNARFALNAANRWGSLDYALYGTQVIVPETDGAEGKPTYNKVRGDKVIAYAR 180
Db 126 GPQLVPLVNARFALNAANRWGSLDYALYGTQVIVPETDGAEGKPTYNKVRGDKVIAYAR 185
Qy 181 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSLLI 240
Db 186 KFLDDSVPLSSGSGFGDATGFTVODGOLVVALPKSTGLANPGQFAGYTGAAESPTSLLI 245
Qy 241 NHGLHIEILIDPSQVGTTRAGVKDVIIESAITTIMDFEDSVAADAAKVLGYRNWLG 300
Db 246 NHGLHIEILIDPSQVGTTRAGVKDVIIESAITTIMDFEDSVAADAAKVLGYRNWLG 305
Qy 301 LNKGDIAAAVDKGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIIVDT 360
Db 306 LNKGDIAAAVDKGTAFRLVNRDRNYTAPGGQFTLPGRSLMFVRNVGHLMTNDIIVDT 365
Qy 361 DGSEVPEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 420
Db 366 DGSEVPEGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPMHGPAAVFTCELF 425
Qy 421 SRVEDVLGLPQNTMKIGIMDEERTTNLKCACIKAAADRVPFINTGFLDRTGDEIHTSME 480
Db 426 SRVEDVLGLPQNTMKIGIMDEERTTNLKCACIKAAADRVPFINTGFLDRTGDEIHTSME 485
Qy 481 AGPMVRKGTMTKSPWILAYEDHNVDAAGSRAQVKGKMTMTLMADVMVETKIAQP 540
Db 486 AGPMVRKGTMTKSPWILAYEDHNVDAAGSRAQVKGKMTMTLMADVMVETKIAQP 545
Qy 541 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 600
Db 546 RAGASTAWVSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDE 605
Qy 601 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSELLANLWLRHGVI 660
Db 606 IREVDNCCSILGYVVRWDQVGGSKVPDIHDVALMEDRATLRISSELLANLWLRHGVI 665
Qy 661 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPI 720
Db 666 TSADVRSASLERMAPLVDQRNAGDVAVRPMAPNFDDSIAPLAAQELILSGAQQPNGYTEPI 725
Qy 721 LHRRRREFKARAAEKPAAPSDRAGDDAAR 748
Db 726 LHRRRREFKARAAEKPAAPSDRAGDDAAR 753
```

RESULT 3

```
US-09-738-626-6988
; Sequence 6988, Application US/09738626
; Publication No. US20020197605A1
```

GENERAL INFORMATION:

```
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6988
```



```

; LENGTH: 739
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6988

Query Match      37.8%; Score 2289; DB 10; Length 739;
Best Local Similarity 61.3%; Pred. No. 2.2e-188;
Matches 444; Conservative .90; Mismatches 184; Indels 6; Gaps 4;

QY 9 TDRVSGNLRARVLYDFVNEALPGTDIDPDSFWAGVDKVVADLTTPQNALNARELQ 68
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 20 TERVDAGMVAKVLYDFVTEAVLPVGVDAKFWGFAAIARDLTFRNELLARRDELQ 79

QY 69 AQIDKWHRRVIEPIDMDAYEQFITEIGYLLPEPDDFTTITSGVDABITTTAGQLVVPV 128
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 MLIDYHRNN-SGITDQEAIVEDFKEIGYLVVEPEAAEIRTONVDTEISISSTAGQLVVPV 138

QY 129 LNARFALNAARWGSVLDALYDGTVDIPETDGAEGKPTYNKVRGDKVIAYARFELDQVSP 188
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 LNARFALNAARWGSVLDALYGNATIPETDGAEGKEYNPVRQKVIEWGREFLDSVVP 198

QY 189 LSSGSFGDATGFTVQDQGVVALPKSTGLANPQCFAGYTGAAESP*SVLLIHNGLHIEI 248
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 LDGASHADVEKYNITDGLAAHIGDSVYRLKNRESYRGFTGNFLDPEAILLETNGLHIEL 258

QY 249 LIDPESQVTTDRAGVKDVILLESATITIMDPEDSVAADKVLGYRNWGLNKGDLAA 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 QIDPVHPITGKADKTGLKDIVLESATITIMDPEDSVAADKLTGYSNFWGLNTGELKE 318

QY 309 AVDKDGTAFRLVNRDRNTYAPGGQFTLPGRSLMFVRNVGHLMTNDIAIVDTDQSEVFEG 368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 EMSXNGRIFTRELANKDRVYIGENGTELVLHGSSLLFVRNVGHLMTNPSIL-IDGEEIFEG 377

QY 369 IMDALFTGLIAIHGLKASDVNGPLINRSTGSIYIVKPKMHGPAVFTCELFSRVEDVLG 428
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 IMDAVLTTCVCAIPGAPQN---KMRNSRKSGSIYIVKPKQHGEVAVFTNELFGREVDLLD 434

QY 429 LPQNTMTKIGWDEERRTTVNLKACIKAAADRVVINTGFLDRTGDETHTSMEAGPMYRKG 488
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 435 LPRHLKVGWDEERRTSVNDJASIMEVADRLATINTGFLDRTGDETHTSMEAGMYRKA 494

QY 489 TMKSQPTILAYEDHNVDAAGLAFSGRAQVKGKMTMTLMADMVETKIAQPRAGASTAW 548
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 495 DMQTAQKQAYENNVNDAGIQRLPGKAQIKGKMWMTLMMAEMLEKKIQPREGANTAW 554

QY 549 VPSPTAATLHALYHQVDVAAVQOGL-AGKRATIEQLITITPLAKELAWAPDEIREVDN 607
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 555 VPSPTGATLHATHVHLVDVFKVQDELRAAGRRDSLRLNITITPTAPNTNWSEEEKKEEMDN 614

QY 608 NCQSILGYVVRWVQGVCSKVPIHDHVALMEDRATLRISSQLLANWLRHGVITSADVRA 667
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 615 NCQSILGYVVRWVHGVCCKVPIHDHIDLMDRATLRISSQMLANWIRHDVVSKEQVLE 674

QY 668 SLERMAPLVDRQAGDVAYRMAPNFDDSIATFLAAQELILSGAQQPNGYTEPILHRRRE 727
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 675 SLERMAVVDRQAGDEAYRMAPNYDASLAFQAADLIFEGTKSPSGYTEPILHARRRE 734

QY 728 FKAR 731
   |||:|||||:
Db 735 FKAK 738

RESULT 4
US-09-793-306-163
; Sequence 163, Application US/09793306
; Patent No. US20020058200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shvian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation

Query Match      26.7%; Score 1615.5; DB 9; Length 421;
Best Local Similarity 85.1%; Pred. No. 1.3e-130;
Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;

QY 749 VKYGGSSVADAEIRRVAAERIVATKKQGNVVVVVSAMGDTTDDLDDLAQQVCAPPPR 808
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 VKYGGSSVADAEIRRVAAERIVATKKQGNVVVVVSAMGDTTDDLDDLAQQVCAPPPR 64

QY 809 ELDMLLTAGER:SNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKI:IDVTPGRLOTA 868
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ELDMLLTAGER:SNALVAMAIESLGAHARSFTGSGAGVITTTGTHGNAKI:IDVTPGRLOTA 124

QY 869 LEEGRVVLVAGQGVSDTKDVTTLGRGSGDTTAVAMAALGADVCEIYTDVGDGIFSADP 928
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 LEEGRVVLVAGQGVSDTKDVTTLGRGSGDTTAVAMAALGADVCEIYTDVGDGIFSADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLRCEVYARRHNI:PVHVRSSYSYSDRPGTVVVG 988
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 RIVRNARKLDTVTFEEMLEMAACGAKVLMRLRCEVYARRHNI:PVHVRSSYSYSDRPGTVVVG 244

QY 989 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIFGYAAKVFRAVARRRRQRHGAEE----- 1043
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 IKDVPMDPILTVGAHDSRSEAKVTIVGLPDIFGYAAKVFRAVARRRRQRHGAEE-----DADVNI 295

QY 1044 RLQGRQRQRHHLHLLPQTS---GPPPKQWNRSETRSRSTOLLYDDHIGKVSILGAGMR 1100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 296 VLQNVSKVEDGKTDITFTCSRVDVGAAPAEKLSLRNEIGFSQLLYDDHIGKVSILGAGMR 355

QY 1101 SHPGVTATFCEALAAVGVNIELISTSESDQSRCCAAAT 1137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 356 SHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDT 392

RESULT 5
US-09-738-626-3781
; Sequence 3781, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, WASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3781
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3781

Query Match      20.2%; Score 1224.5; DB 10; Length 421;
Best Local Similarity 64.0%; Pred. No. 7.7e-97;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABRIRRVABRIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPPPPR 808
DB 5 VQYGGSSLESABRIRRVABRIVATKKAGNDVVVVCAMGDTTDELLELAAAVNPVPPAR 64

QY 809 ELDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNARIVDVTPGRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAAAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTVVVGVS 988
DB 185 RIVNARKLEKUSFEEMLEMAAGSKILVRSVEYARAFNPVRVRSYSDNPGTLLIAGS 244

QY 989 IKDVPNEDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
DB 245 MEDIPVEEAVLTGVTADTKSEAKVTVLGSDKPGEAAKVFRALADAEINIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPFPKWNTRSETRSASTQLLYDDHIGKVSLLI 1095
DB 305 DGTDTDTFTCPRSDGRAMEILKLVQVG-----TNVLYDDQVGKVSLLV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

RESULT 6
US-09-746-660A-24
; Sequence 24, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CP2
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970

; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 24
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-746-660A-24

Query Match      20.2%; Score 1224.5; DB 11; Length 421;
Best Local Similarity 64.0%; Pred. No. 7.7e-97;
Matches 251; Conservative 55; Mismatches 59; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABRIRRVABRIVATKKQGNVTVVVSAMGDTTDDLDAQQVCPPPPR 808
DB 5 VQYGGSSLESABRIRRVABRIVATKKAGNDVVVVCAMGDTTDELLELAAAVNPVPPAR 64

QY 809 ELDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNAKIIDVTPGRLOTA 868
DB 65 EMDMLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTGTHGNARIVDVTPGRVREA 124

QY 869 LEEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAAAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVNARKLDTVTPEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTVVVGVS 988
DB 185 RIVNARKLEKUSFEEMLEMAAGSKILVRSVEYARAFNPVRVRSYSDNPGTLLIAGS 244

QY 989 IKDVPNEDPILTVGAHDSRSEAKVTIVGLPDIPGYAAKVFRAVA-----RRRRQHR 1038
DB 245 MEDIPVEEAVLTGVTADTKSEAKVTVLGSDKPGEAAKVFRALADAEINIDMVLQNVSSVE 304

QY 1039 HGAERLQGRQD-RHHLHLP--QTSPPFPKWNTRSETRSASTQLLYDDHIGKVSLLI 1095
DB 305 DGTDTDTFTCPRSDGRAMEILKLVQVG-----TNVLYDDQVGKVSLLV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127
DB 351 GAGMKSHPGVTAEFMEALRDVNVNIELISTSE 382

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/226,136
```

```

; ORGANISM: Corynebacterium glutamicum
US-10-067-974-2

Query Match          20.1%; Score 1217.5; DB 15; Length 421;
Best Local Similarity 63.8%; Pred. No. 3.1e-96;
Matches 250; Conservative 55; Mismatches 60; Indels 27; Gaps 5;

QY 749 VQYKGSSVADAEIRIRVAERIIVATKKQGNVVVVVSAMGDTTDDLLDLAQCYCPAPPPR 808
Db 5 VQYKGSSVLESAEIRIRVAERIIVATKKAGNDVVVVCAMGDTTDELELAAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVMAAIESLGAHARSFTGSGAQVITGTGHNAKIIDVTPGRLQTA 868
Db 65 EMDMLLTAGERISNALVMAAIESLGAFAQSFTGSGAGVLTERHGNARIYDVTPGVRVREA 124

QY 869 LEBGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAVAAALGADVCEIYTDVIGIFISADP 928
Db 125 LDEGKICIVAGFGVQNKETRDVTTLGRGSGDTTAVAAALNADVCEIYSDVDGVYTADP 184

QY 929 RIVRNAKGLDVTTFEEMLEMAACGAKVLMRLCVEYARRHNI PVHVRSSYSDRPCTVVVGS 988
Db 185 RIVPNAQKLEKLSFEEMLELAAGVSKILVRSVEYARAFNVPRVRSSYSNDPGLTIAGS 244

QY 989 IKDVPMDPILTCVADHRSEAKVTIIVGLPDPGYAAKVFRAVA-----RRRQHR 1038
Db 245 MEDIPVEEAVLTVATDKSEAKVTILGISDKPGEAAKVFALADAETINIDWLQNVSSE 304

QY 1039 HGAERLQGRGCD-RHLLHLP--QTSQPPPKWTRSETRASTOLLVDDHIGKVS LI 1095
Db 305 DGTDTITFCPRADGRAMEILKXLQVG-----TNVLXDDQVGKVS LV 350

QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
Db 351 GAGMKSHPGVTAFEMEALRDVNVNIELISISE 382

RESULT 9
US-10-067-974-16
; Sequence 16, Application US/10067974
; Publication No. US2003005232A1
; GENERAL INFORMATION:
; APPLICANT: Li, Liang-Yew
; APPLICANT: Trei, Kelli J.
; TITLE OF INVENTION: Polynucleotide Constructs for Increased Lysine Production
; FILE REFERENCE: 1533.2640001
; CURRENT APPLICATION NUMBER: US/10/067,974
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,183
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus Sequence of Protein Sequence Alignment
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (40)..(40)
; OTHER INFORMATION: May be either Cys or Val
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (317)..(317)
; OTHER INFORMATION: May be either Ser or Ala
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: May be either Gly or Asp
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (380)..(380)
; OTHER INFORMATION: May be either Thr or Ile

```

US-10-067-974-16

Query Match 20.0%; Score 1211.5; DB 15; Length 421;
 Best Local Similarity 63.5%; Pred. No. 1e-95;
 Matches 249; Conservative 55; Mismatches 61; Indels 27; Gaps 5;

QY 749 VQYGGSSVADABIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQVCPAPPPR 808
 Db 5 VQYGGSSLESABIRIRVAERIVATKKAGNVVWVVSAMGDTTDELLELAANVPVPPAR 64
 QY 809 ELDMLLTAGERISNALVVAISLGHARSFTGSQAGVITTGTHGNAKIIDVTPGRLOTA 868
 Db 65 ELDMLLTAGERISNALVVAISLGHARSFTGSQAGVITTEHGNARIVDVTGRVREA 124
 QY 869 LEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
 Db 125 LDEGKICIVAGFGVQVNETRDVTTLGRGSGDTTAVAMAAALNADVCEIYSDVDGVYTADP 184
 QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 988
 Db 185 RIVNARKLEKLSFEEMLEMAAGSKILVRSVEYARRHNPVHVRSSYSNDRPGTILAGS 244
 QY 989 IKDVPMPDILTGVAHDSRSEAKVTIYGLPDIPGYAAKVFRAVA--RRRQHR 1038
 Db 245 MEDIPVEEAVLTGATDKSEAKVTVLGSDKPCGEAAKVFRAADAEINIDVWLQNVSSVE 304
 QY 1039 HGAERLQGRGROD-RHLLHLLP--QTSGPPPKWTRSETRSASTOLLYDDHIGKVSLLI 1095
 Db 305 DGTDTITFTCPRXDGRAMEILKQLVQG-----TNVLYDDQVXKSLV 350
 QY 1096 GAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGMKSHPGVTAEFMEALRDVNVNIELISXSE 382

RESULT 10

US-10-156-761-12093
 ; Sequence 12093, Application US/10156761
 ; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 12093

LENGTH: 430

TYPE: PRT

ORGANISM: Streptomyces avermitilis

US-10-156-761-12093

Query Match 18.9%; Score 1144.5; DB 15; Length 430;
 Best Local Similarity 60.2%; Pred. No. 6.5e-90;
 Matches 239; Conservative 54; Mismatches 75; Indels 29; Gaps 4;

QY 749 VQYGGSSVADABIRIRVAERIVATKKQGNVWVVSAMGDTTDDLDAQVCPAPPPR 808
 Db 5 VQYGGSSVADABIRIRVAERIVATKKAGNVVWVVSAMGDTTDELLELAANVPVPPAR 64
 QY 809 ELDMLLTAGERISNALVVAISLGHARSFTGSQAGVITTGTHGNAKIIDVTPGRLOTA 868
 Db 65 ELDMLLTAGERISNALVVAISLGHARSFTGSQAGVITSDVHKNARIIDVTPGRIRTA 124

QY 869 LEGRVVLVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGIFSADP 928
 Db 125 LDEGNIIVAGFGVQSDTKDVTTLGRGSGDTTAVAMAAALDAEVCEIYTDVDFGVTADP 184
 QY 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSRDRPGTVVVG 988
 Db 185 RWKAKKIDWIAFEDMLELAASGSKVLLRCEYARRYNIPVHVRSSFSGLQGTWV--- 241
 QY 989 IKDVP-----EDPILTGVAHDSRSEAKVTIYGLPDIPGYAAKVFRAVAARR-- 1034
 Db 242 -SNTFLVQKQQGQEQVQAIIISGVVADTSEAKVTIYGLPDIPGYAAKVFRAVDAEYN 300
 QY 1035 ----RQRHGAERLQGRGROD-RHLLHLLP-QTSGPPPKWTRSETRSASTOLLYDDHIG 1090
 Db 301 IDMVVQNVSAASTGLTDIS-----FTLPKTEGRKAIDALEKASVIGFDSLYDDQIG 353
 QY 1091 KVSLLGAGMRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 354 KISLVGAGMKTNPGVTAGTFFEALSDAGVNIELISTSE 390

RESULT 11

US-10-084-843-227
 ; Sequence 227, Application US/10084843
 ; Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, David C.
 Campos-Neto, Antonio
 Houghton, Raymond
 Vedwick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843

FILING DATE: 25-Feb-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 227:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 227:

US-10-084-843-227

Query Match 12.7%; Score 770; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.7e-58;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 750 QYGGSSVADAEIRIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRE 809
Db 1 QYGGSSVADAEIRIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRE 60

Qy 810 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIVTPGRLOTAL 869
Db 61 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIVTPGRLOTAL 120

Qy 870 BEGRVVLVAGFQGVSDTKDVTTLGRGGSDDTTAVAM 905
Db 121 BEGRVVLVAGFQGVSDTKDVTTLGRGGSDDTTAVAM 156

RESULT 12

US-10-193-002-222
; Sequence 222, Application US/10193002
; Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002

FILING DATE: 10-Jul-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 222:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 222:

US-10-193-002-222

Query Match 12.7%; Score 770; DB 12; Length 156;

Best Local Similarity 100.0%; Pred. No. 2.7e-58;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 750 QYGGSSVADAEIRIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRE 809
Db 1 QYGGSSVADAEIRIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRE 60

Qy 810 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIVTPGRLOTAL 869
Db 61 LDMLLTAGERISNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIVTPGRLOTAL 120

Qy 870 BEGRVVLVAGFQGVSDTKDVTTLGRGGSDDTTAVAM 905
Db 121 BEGRVVLVAGFQGVSDTKDVTTLGRGGSDDTTAVAM 156

RESULT 13

US-09-882-227-290

; Sequence 290, Application US/09882227

; Publication No. US20030159396A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Coomen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20030159396A1 Helicobacter Polypeptides in the

FILE REFERENCE: 06132/047002

CURRENT APPLICATION NUMBER: US/09/882,227

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/902,615

PRIOR FILING DATE: 1997-07-29

NUMBER OF SEQ ID NOS: 638

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 290

LENGTH: 394

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-882-227-290

Query Match 11.1%; Score 673; DB 12; Length 394;

Best Local Similarity 40.1%; Pred. No. 3e-49;

Matches 151; Conservative 83; Mismatches 119; Indels 24; Gaps 6;

Qy 761 ERIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRELDMLTAGERI 820

Db 5 ERIRVAERIVATKQGNQNDVVVVVSAMGDTTDDLDAQQVCPPPRELDMLTAGERI 64

Qy 821 SNALVAMAEISLGAHARSFTGSGAGVITTTGTHGNAKIIVTPGRLOTALBEGRVVLVAGF 880

Db 65 SSALSNALERYGHRALSLSGKEAGILTSCHFQNAVIOISIDTKRITELLEKNYVVIAGF 124

Qy 881 QGVSDTKDVTTLGRGGSDDTTAVAMAAALGADVCEIYTDVDFISADPRIVRNARKLDTV 940

Db 125 QG-ADIQGETTLGRGGSDDLAVAGALKAHLCIEIYTDVGVYTTDPRIEERAKIAQI 183

Qy 941 TFEEMLEMAACGAKVLMRCVEYARRHNIPVHVSSYSDRPGTVAWG--SIKQVPMEDPI 998

Db 184 SYDEMELASGAKVLLNRSVLAKLSVKLVTRNSFNHSEGLIIVAEKDFKERNETPI 243

Qy 999 LTGVADHRSKAKYTIIVGLPIPGVAAKVFRAVARRRRQHRHGAERLQGRGRDRHHLHL 1058

Db 244 VSGIALDNQARVSMEGVEDRPGIAABIFGALA---EYRINVDIMVQTTGRDKTDLDF 299

Qy 1059 -LPQTSPPPKWTRSRSTRSASTOLL-----YDDHTGKVLISLGAGMPSHQVTATFC 1110

Db 300 TIVTKT-----IETKQALKPFLAQMDSIDYENTAKVSIIVGVGMKSHSGVASIAF 350

Qy 1111 EALAAVGVNIELSTSE 1127

Db 351 KALAKDNINIMISTSE 367

RESULT 14

US-09-890-813-17
; Sequence 17, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/172944
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Glycine max
US-09-890-813-17

Query Match 6.2%; Score 377; DB 10; Length 564;

Best Local Similarity 27.0%; Pred. No. 2e-23;
Matches 128; Conservative 71; Mismatches 165; Indels 110; Gaps 13;

QY	749	VQYGGSSVADAEIRIRVAERIVATKKQGNVVDVVVVSAMGDTTDDL	-----	795
Db	87	VMFEGSSVASADRMKEVATLILSPFEE--RPVIVLSANGKTNKLLAGEKAVSCGVIN	144	
QY	796	-----DLAQVCPAPPREL	-----	817
Db	145	VSSIEELCFIKDHLRTVDQLGVDSVISKHLEELBQLKGIAMMKELTKRTQDYIVSFG	204	
QY	818	ERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVT---PGRLO--TALBEG	872	
Db	205	ECMSTRIFRAYLNKIGVKARQYDAFEGITTTDFNADILEATYPVAKRLHGDWLSDP	264	
QY	873	RVVLVAGFGVSGDTKDVTLGRGSDTTAVAMAALGADVCEIYTDVGIFSDPRIVR	932	
Db	265	AIATVTGFLGKARKSAVITTLGRGSDLTATTIGKALGPEIQVWKDVGVLTCDPNIYP	324	
QY	933	NARKLTVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSR--PGTVVGSIKD	991	
Db	325	KABPVPLTFDEAAELAYFGAQLHFCQMRPARESIPVRVKNVSNPKAPGFLIT---KA	381	
QY	992	VPMEDPILTVGAHDRSEAKVTIVGLPDIP--GYAAKVFRAVARRRRQHRHGAERLOGRG	1049	
Db	382	RDMSKAVLTSIVLKRNVMTLDIASTRMLGQYGLAKVFSIFEELGISVDVATSEVSYS	441	
QY	1050	RDRHHLHLLPQTSGPPKWNTRSTRSASTQLLYDDHIGK-----VS	1093	
Db	442	TLD-----PSKJMSRELIOQASEL---DHVVEELKIAVNVNLLQNRSTIS	483	
QY	1094	LIGAGMRSHPGVATPCEALAAVGNVIELIS-----TSEDQSRCCAA	1136	
Db	484	LIGNVQRSSL--ILERSLVRLTIGTVQMSIQASKVNIISLVVNDSEAEQCVRA	536	

RESULT 15

US-09-890-813-8
; Sequence 8, Application US/09890813
; Publication No. US20020183486A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Aspartate Kinase
; FILE REFERENCE: BB1430 PCT
; CURRENT APPLICATION NUMBER: US/09/890,813
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/172944
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 8
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Zea mays
US-09-890-813-8

Query Match 6.0%; Score 365; DB 10; Length 555;

Best Local Similarity 29.9%; Pred. No. 2.1e-22;
Matches 103; Conservative 58; Mismatches 119; Indels 64; Gaps 7;

QY	743	GDDAARQYKYGSSVADAEIRIRVAERIVATKKQGNVVDVVVVSAMGDTTD	-----	792
Db	80	GDQLSVVMKFGSSVSSAARMAEVAGLILTFPEE--RPVVVLSAMGKTTNNLLAGEKAV	137	
QY	793	-----DLDLAQVCPA	-----	811
Db	138	CGVIVHSEIEEEMWKSLLHKTVDLGLPRSVIQDMLDEQLLKGIAWMKELTPRTSD	197	
QY	812	MLLTAGERISNALVAMAIESLGAHARSFTGSOAGVITTTGTHGNAKIIDVT	869	
Db	198	YLVSFGECSMSTRIFSAVNLKIRVKARQYDAFIDGFIITDFEGNADILEATYPVAKRLHG	257	
QY	870	---BGRVVLVAGFGVSGDTKDVTLGRGSDTTAVAMAALGADVCEIYTDVGIFSA	926	
Db	258	DWIQDPAIPVVTGFLGKWSGAVTTLGRGSDLTATTIGKALGRLREIQWKDVGVLTC	317	
QY	927	DPRIVRVARKLDVTTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSR--PGTVV	985	
Db	318	DPNIYPHAKTVPYLTTFEATELAYFGAQLHFCQMRPAREGDIPVRVKNVSNPKAPGTLI	377	
QY	986	VGSIKDVPMBDPIITGVGAHDRSEAKVTIVGLPDIP--GYAAKVF	1027	
Db	378	T---RQRDMDKWLTSTVLKSNVTMLDIVSTRMLGQYGLARVF	418	

Search completed: November 21, 2003, 16:38:14

Job time : 46.7107 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 24.4499 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-52
Perfect score: 6061
Sequence: 1 MQHHHHHTDRVSVGNLRIA.....SAATRRPCTGRGRWACQ 1172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3810	62.9	741	2 F70722	probable glcB prot
2	3117.5	51.4	731	2 T44752	probable malate sy
3	2569	42.4	725	2 H83586	malate synthase G
4	2380	39.3	744	2 AH2582	malate synthase G
5	2380	39.3	744	2 F97364	malate synthase G
6	2318	38.2	728	2 AP3299	malate synthase (E
7	2289	37.8	739	2 I40715	malate synthase (E
8	2258	37.3	727	2 E83916	malate synthase BH
9	2154.5	35.5	723	2 S51788	malate synthase (E
10	1615.5	26.7	421	2 F70794	probable ask prote
11	1545.5	25.5	421	2 G87199	aspartokinase (imp
12	1478.5	24.4	421	2 S42422	aspartate kinase
13	1229.5	20.3	421	2 S15276	aspartate kinase
14	1141	18.8	425	2 T33383	probable aspartoki
15	823	13.6	412	2 C83531	aspartate kinase a
16	816.5	13.5	600	2 S76764	hypothetical prote
17	785	13.0	411	2 A48946	aspartate kinase
18	778	12.8	407	2 JC4640	aspartate kinase
19	776.5	12.8	606	2 AE2261	aspartate kinase
20	761	12.6	405	2 H81865	aspartate kinase
21	760	12.5	405	2 F81076	aspartokinase, alp
22	757.5	12.5	412	2 H84036	aspartokinase II a
23	742.5	12.3	400	2 E81405	aspartate kinase
24	736.5	12.2	739	2 H72364	aspartokinase II -
25	734.5	12.1	415	2 B70399	aspartokinase - Aq
26	724.5	12.0	401	2 G72245	aspartokinase II -
27	721	11.9	405	2 H71843	aspartokinase 2 al
28	712.5	11.8	408	2 A29314	aspartate kinase
29	711	11.7	405	2 E64673	aspartokinase - He

RESULT 1

F70722

Probable glcB protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: F70722

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70722

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-741 <COL>

A;Cross-references: GB:278020; GB:AL123456; NID:g3261625; PIDN:CAB01465.1; PID:e257679;

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: glcB

Query Match 62.9%; Score 3810; DB 2; Length 741;
Best Local Similarity 100.0%; Pred. No. 7.1e-221; Indels 0; Gaps 0;
Matches 740; Conservative 0; Mismatches 0

Qy 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVQKXVADLTPOQALLNARDELQ 68

Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFWAGVQKXVADLTPOQALLNARDELQ 61

Qy 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPPDDFTITTSVDABITTTAGPOLVVPV 128

Db 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPPDDFTITTSVDABITTTAGPOLVVPV 121

Qy 129 LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAYARFLODSVP 188

Db 122 LNARFALNANARWGLYDALYGTDPETDGAEGKPTYNKVRGDKVIAYARFLODSVP 181

Qy 189 LSSGSFGDAGFTVQDQQLVVALPDKSTGLANPQGFAGYTGAAESPTSVLLINHLHIEI 248

Db 182 LSSGSFGDAGFTVQDQQLVVALPDKSTGLANPQGFAGYTGAAESPTSVLLINHLHIEI 241

Qy 249 LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADKVLGYRNWGLNKGDLAA 308

Db 242 LIDPESQVGTTRAGVKDVILESAITTIMDFEDSVAADKVLGYRNWGLNKGDLAA 301

Qy 309 AVDKDGTAFRLVLRNDRNTAPGGQQTFLPGRSLMFVRNVGHMLTNDIAIYDTDGSEVFEG 368

Db 302 AVDKDGTAFRLVLRNDRNTAPGGQQTFLPGRSLMFVRNVGHMLTNDIAIYDTDGSEVFEG 361

Qy 369 IMDALFTGLTAIRHCLKASDVNGPLINSRTGSIIVVRKMGHPAEVATCFLFSRVEDVLG 428

Db 369 IMDALFTGLTAIRHCLKASDVNGPLINSRTGSIIVVRKMGHPAEVATCFLFSRVEDVLG 428

Db 362 IMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAVFAFTCELSRV 421
QY 429 LPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRGTGDEIHTSMEAGPVRK 488
Db 422 LPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRGTGDEIHTSMEAGPVRK 481
QY 489 TMSQSPWILAYEDHNVDAGLAAGFSGRAQVKGKMTMELMADVETKIAQPRAGASTAW 548
Db 482 TMSQSPWILAYEDHNVDAGLAAGFSGRAQVKGKMTMELMADVETKIAQPRAGASTAW 541
QY 549 VPSTPAATHALHYHQVDVAOVQGLAGKRRATIEQLLTIPLAKELAWAPDEIREVDNN 608
Db 542 VPSTPAATHALHYHQVDVAOVQGLAGKRRATIEQLLTIPLAKELAWAPDEIREVDNN 601
QY 609 CQSTILGVVVRWVQGVGCSKVPDIHDVLMEDRATIRISSQLLANWLHGVITTSADVRAS 668
Db 602 CQSTILGVVVRWVQGVGCSKVPDIHDVLMEDRATIRISSQLLANWLHGVITTSADVRAS 661
QY 669 LERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHRRREF 728
Db 662 LERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHRRREF 721
QY 729 KARAEEKPAPSDRAGDDAAR 748
Db 722 KARAEEKPAPSDRAGDDAAR 741

RESULT 2
T44752
probable malate synthase (EC 4.1.3.2) G [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
C:Accession: T44752
R:Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z22833
A:Accession: T44752
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-731 <P>
A:Cross-references: EMBL:AL008609; PIDN:CAA15459.1
A:Experimental source: cosmid B1788
C:Genetics:
A:Note: glcB
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 51.4%; Score 3117.5; DB 2; Length 731;
Best Local Similarity 82.0%; Pred. No. 2.7e-179;
Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;
QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKAVADLTPOQALLNARDDELQ 68
Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKAVADLTPOQALLNARDDELQ 61
QY 69 AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDFTITTSQVDAEITTTAGPQLVVPV 128
Db 62 AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDFTITTSQVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANARWGLYDALYGTVDVTPETDGAEGKPTYNKVGDKVIAYARFLLDDSV 188
Db 122 LNARFALNANARWGLYDALYGTVDVTPETDGAEGKPTYNKVGDKVIAYARFLLDDSV 181
QY 189 LSSGSGFDATGFTVQDQQLVVALPDKSTGLANPGOPAGYTGAAESPT-SVLLINHLGHE 247
Db 182 LASDSWTNATGVSIFDQQLQIAIGTNTSTGLASPEKPVGNRQLRSSNWSVLLANHLGHE 241
QY 248 ILIDPESQVGTTRAGVKDVIKESATITIMDFEDSVAADADKVLGYRNWGLNKGDLA 307
Db 242 VLIDPESQVGTTRAGVKDVIKESATITIMDFEDSVAADADKVLGYRNWGLNKGDLA 301
QY 308 AAVDKDGTAFRLVINDRNVYTPGGQQTLPGRSLMFRVNVGHMTNDAIV----DTGGS 363
Db 302 EEWKDKGKFTFRLVINDRNVYTPGGQQTLPGRSLMFRVNVGHMTNDAIV----DTGGS 361

QY 364 EVFSGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAVFAFTCELSRV 423
Db 362 EVFSGIMDALFTGLIAIHGLKASDVNGPLINSRTGSIYIVKPKHGPAAVFAFTCELSRV 421
QY 424 EDVLGLPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRGTGDEIHTSMEAGP 483
Db 422 EDVLGLPQNTMKIGIMDEERTTNLTKACIKAAADRVVFINTEGFLDRGTGDEIHTSMEAGP 481
QY 484 MYRGTMTKSPWILAYEDHNVDAGLAAGFSGRAQVKGKMTMELMADVETKIAQPRAG 543
Db 482 MIRKAMKNSWILAYEDHNVDAGLAAGFSGRAQVKGKMTMELMADVETKIAQPRAG 541
QY 544 ASTAWVSPPTAATHALHYHQVDVAOVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRE 603
Db 542 ASTAWVSPPTAATHALHYHQVDVAOVQGLAGKRRATIEQLLTIPLAKELAWAPDEIRE 601
QY 604 EVDNQCOSILGVVVRWVQGVGCSKVPDIHDVLMEDRATIRISSQLLANWLHGVITTS 663
Db 602 EVDNQCOSILGVVVRWVQGVGCSKVPDIHDVLMEDRATIRISSQLLANWLHGVITTS 661
QY 664 DVRSASLERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHR 723
Db 662 DVRSASLERMAPLVDRQNGADVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYTEPILHR 721
QY 724 RRREFKAR 731
Db 722 RRREFKAR 729

RESULT 3

H83586
malate synthase G PA0482 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83586
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brice, S.; Young, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83586
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <STO>
A:Cross-references: GB:AF004485; GB:AE004091; NID:g99946332; PIDN:AAG03871.1; GSPDB:GN0013
A:Experimental source: strain PA01
C:Genetics:
A:Gene: glcB; PA0482

Query Match 42.4%; Score 2569; DB 2; Length 725;
Best Local Similarity 69.1%; Pred. No. 2.3e-146;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
QY 9 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKAVADLTPOQALLNARDDELQ 68
Db 2 TDRVSVGNLRIARVLYDFVNNALPGTDIDPDSFAGVGVKAVADLTPOQALLNARDDELQ 61
QY 69 AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDFTITTSQVDAEITTTAGPQLVVPV 128
Db 62 AQIDKWHRRVIBPIDMDAYRQFLTEIGYLLPEPDFTITTSQVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANARWGLYDALYGTVDVTPETDGAEGKPTYNKVGDKVIAYARFLLDDSV 188
Db 122 LNARFALNANARWGLYDALYGTVDVTPETDGAEGKPTYNKVGDKVIAYARFLLDDSV 181
QY 189 LSSGSGFDATGFTVQDQQLVVALPDKS-TGLANPGOPAGYTGAAESPTSVLLINHLGHE 247
Db 182 LSSGSHVDATSYSVKNGALVVALKNGSETGLKNGAFLAQFGDAAPQAVLLKHGLHFE 241
QY 248 ILIDPESQVGTTRAGVKDVIKESATITIMDFEDSVAADADKVLGYRNWGLNKGDLA 307


```
Db 242 IQIDSSPVGQDAAAGVKVLMKALTTIMDCEDSVAADVADDDKVIYRNWGLMKGDLA 301
QY 308 AAVDKDGTAFURVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHMTNDIAIVDTGSEVFE 367
Db 302 EVSFGGSGTFTRTMNPDRVYTRADGSELTLHGRSLFVRNVGHMTNDIAILDKDCNEVPE 361
QY 368 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKMHGPAEVAFTCELFSEVED 425
Db 362 GJQDGLFTSLIAIH-----DLNGNTRKNSRTGSIYIVKPKMHGPEEAFTNEIFGRVED 416
QY 426 VLGLPQNTWKIGIMDEERTTNVNLKACIKAAADRVPVINTGFLDRTGDEIHTSMEAGPMV 485
Db 417 VLGLPRLNTLKVIMDEERTTNVNLKACIKAAAKDRVPVINTGFLDRTGDEIHTSMEAGAVV 476
QY 486 RKGTKWKSQPTLAIYEDHNVDAAGLAGFSGRAQVKGKMTMELVADNVETKIAOPRAGAS 545
Db 477 RKGAKSKKWGAYENNNDVGLATGLOKQAIQIGKMWAMPDLMAAMLEQKIGHPLAGAN 536
QY 546 TAWPSPATAATHALHYHQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIRBEV 605
Db 537 TAWPSPATAATHALHYHKVDVFAQAELAKRTPASVDDILTIPLAPNTNTWASEIKNEV 596
QY 606 DNNCOSILGYVVRVWDQGVGSKVPDIHDVAMDEBRATLRISQQLANLWLRHGVITSADV 665
Db 597 DNNAGILGYVVRVWDQGVGSKVPDIHDVGLMEDRATLRISQQLANLWLRHGVISQBV 656
QY 666 RASLERMAPLVDRONAGDVAVRPMAPNFDSDTAFLAAOELILSGAQONGYTEPIHLHRR 725
Db 657 VESLXMAVVDQRNASDPSVRPMAPNFDNDVAFQAELVVEGTROQNGYTEPIVLRHR 716
QY 726 REFKA 731
Db 717 REFKA 722

RESULT 4
AH2582
malate synthase G [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2582
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2582
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41078.1; PID:gi7738367; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glcB
A:Map position: circular chromosome

Query Match 39.3%; Score 2380; DB 2; Length 744;
Best Local Similarity 63.5%; Pred. No. 5.3e-135;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;

QY 8 HTRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNALNARDEL 67
Db 22 HVSRTDKFGLSIDRLYAFLTDEVLPCTGLDSTFFEGSAIVHLSKPNRELLAKRDAL 81
QY 68 QAQIDKWHRRVRIEPIMDAYRQFTLGYLLPEPDDTTITSGVDAITTAGPQLVVP 127
Db 82 QKIDGWRENG-ARSDFDAYEAFKEIGYLLPEGPGFKEVNNVDPPIAVVAGPQLVVP 140
QY 128 VLNAPALNAANARWGSLYDALYGTDAISDADGAEEKRGYNPKRGDKVIAWENFLDESA 200
```

```
141 VMNARVALNAANARWGSLYDALYGTDAISDADGAEEKRGYNPKRGDKVIAWENFLDESA 200
QY 188 PLSGSGFGATGTVOGQVVALPKSTGLANPGQFAGYTGAESPTSVLLINHGHLIE 247
Db 201 PLETGWSVDVTGNIADGLLQIAIGAAATGLKDAVOFGKFSGEAKPATILLKNGHLTE 260
QY 248 ILDPSQVQTTDRAGVKDVILLESATTTIMDRDSVAADDAADKVLGYRNWGLNKGDLA 307
Db 261 IVIDPSTEIGKSRAGISDVILLESATTTIMDCEDSVAADDAADKVLGYRNWGLNKGDLT 320
QY 308 AAVDKDGTAFRLVLRNDRNYTAPGGQFTLPGRSLMFVRNVGHMTNDIAIVDTGSEVFE 367
Db 321 EAVSKGNTFTRLNPNDRYYPADGSAITLPGRSLMFVRNVGHMTNPAILDRDGDVPE 380
QY 368 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPKMHGPAEVAFTCELFSEVED 427
Db 381 GIMDAVVITALIYDVGPS---GRRNRSRAGSIYVVKPKMHGPEEVAFAANEIFARVENLV 437
QY 428 GLPQNTWKIGIMDEERTTNVNLKACIKAAADRVPVINTGFLDRTGDEIHTSMEAGPMV 487
Db 438 GMAPNTWKIGIMDEERTTNVNLKESIRAAKDRVPVINTGFLDRTGDEIHTSMEAGPMIRK 497
QY 488 GTWKSQPTLAIYEDHNVDAAGLAGFSGRAQVKGKMTMELVADNVETKIAOPRAGASTA 547
Db 498 GDMKQAATIAAYENWNVDIGECGLSCHAQIGKMWAMPDLMAAMLEQKIAHPKAGANTA 557
QY 548 WPSPTAATLHALHYHQVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPDEIRSEVON 607
Db 558 WPSPTAATLHALHYHKVDVAAVQOGLKSRGAKLSILSVPAVRPNWTPEIQRELND 617
QY 608 NCOSILGYVVRVWDQGVGSKVPDIHDVAMDEBRATLRISQQLANLWLRHGVITSADV 667
Db 618 NAQOILGYVVRVWDQGVGSKVPDINNIGLMEDRATLRISQAQMANLWLRHGVVTEAIIK 677
QY 668 SLERMAPLVDRONAGDVAVRPMAPNFDSDTAFLAAOELILSGAQONGYTEPIHLHRR 727
Db 678 TMKRMARVDTQAGDPAYLPMSDFDGSVAFQAELVVLKGREQNGYTEPIVLRHRL 737
QY 728 FKARAA 733
Db 738 LKAKQA 743

RESULT 5
F97364
malate synthase G (PA0482) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: F97364
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F97364
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85871.1; PID:gl5154912; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 78
A:Map position: circular chromosome

Query Match 39.3%; Score 2380; DB 2; Length 744;
Best Local Similarity 63.5%; Pred. No. 5.3e-135;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;

QY 8 HTRVSVGNLRARVLYDFVNEALPGTDIDPDSFWAGVKVADLTQNALNARDEL 67
Db 22 HVSRTDKFGLSIDRLYAFLTDEVLPCTGLDSTFFEGSAIVHLSKPNRELLAKRDAL 81
QY 68 QAQIDKWHRRVRIEPIMDAYRQFTLGYLLPEPDDTTITSGVDAITTAGPQLVVP 127
Db 82 QKIDGWRENG-ARSDFDAYEAFKEIGYLLPEGPGFKEVNNVDPPIAVVAGPQLVVP 140
QY 128 VLNAPALNAANARWGSLYDALYGTDAISDADGAEEKRGYNPKRGDKVIAWENFLDESA 200
```

Db 82 QEKIDGWYRENG-APSDFDAYEAFLEKIEGYLLPEGPGFKVETNNVDPEIAVAVAGPQLVVP 140
QY 128 VLNARFALNANARWGLSGLDALYGTQVDPETDGAEGKPTYNKVRGDKVIAYARKFLDDSV 187
Db 141 VMARYALNANARWGLSGLDALYGTQVDAISADGAEGKGYNPKRGDKVIAYARKFLDDSV 200
QY 188 PLSSGSGDGTGTVDQGLVVALPKSTGLANPGQFAGYTGAAESPTSVLLINHGHLHIE 247
Db 201 PLETGSDVSTGFIADGLLQALGAATTCGLKDAVQFKGSGEAAKPAITILLKNGHLHIE 260
QY 248 ILDPESQVGTTRDRAGKVDVILSAITTTIMDFEDSVAANDAAKVLGYRNWGLKNGDLA 307
Db 261 IVIDPSTEIGKSRAGISDVILSALTITIMDCDSVAANDAAKVLGYRNWGLKNGDLA 320
QY 308 AAVDKDGTAFRLVLRNDRNTAPGGQFTLPGRSLMFRVNVGHMTNDIAVDTDGSSEVE 367
Db 321 EAVSKGNTTRRLNPDYVYTPDGSALTLPGRSLMFRVNVGHMTNPAILDGRDVE 380
QY 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMHGPAEVAFTCELSRVEDVL 427
Db 381 GIMDAVVTALIALYDVGPS---GRRQNSRAGSMYVVKPKMHGPEEVAFAFANEIPARVENLV 437
QY 428 GLPONTWKIGIMDEERTTNLAKACIAAADRVVFNITGFLDRTGDEIHTSMEAGPMVK 487
Db 438 GMAPNTWKIGIMDEERTTNLAKESRAAKDRVFNITGFLDRTGDEIHTSMEAGPMVK 497
QY 488 GTWKSQWILLAYEDHNVDAAGAFSGRAQVKGKMTMTLMADMVETKIAQPRAGASTA 547
Db 498 GDMKQAIAAYENWNVDIIGLECGLSGHAQIGKGMAMPDMALMALEQKIAHPKAGANTA 557
QY 548 WPSPTAATHALHYHOVDVAAVQOGLAGKRATIEQLITPLAKELAWAPDEIREVDN 607
Db 558 WPSPTAATHALHYHOVDVAAVQOGLAGKRATIEQLITPLAKELAWAPDEIREVDN 617
QY 608 NCOSILGYVVRWVDQGVGCKVPDIHDVVALMEDRATLIRSSQLLANWLRHGVITISADVRA 667
Db 618 NAQILGYVVRWVDQGVGCKVPDIINNGLMEDRATLIRSAQHMANWLRHGVITISADVRA 677
QY 668 SLERMAPLVRQAGDVAYRPMAPNFDSDIAFLAAQELILSGAQONGYTEPILHRRRRE 727
Db 678 TMKMAVVDTONAGDPAYLPNASDFGSAFAAVELVLKQREQNGYTEPILHRRRRE 737
QY 728 FKAAA 733
Db 738 LKAKQA 743

RESULT 6

AF3299
malate synthase (EC 4.1.3.2) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AF3299
R:Delvecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
.; Mazur, M.; Goldsman, E.; Selkov, B.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-728 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51561.1; PID:g17982281; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0380
A:Map position: I
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase
Query Match 38.2%; Score 2318; DB 2; Length 728;
Best Local Similarity 61.8%; Pred. No. 2.7e-131;
Matches 444; Conservative 97; Mismatches 174; Indels 4; Gaps 2;

RESULT 7

I40715
malate synthase (EC 4.1.3.2) - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-May-2000
C:Accession: I40715; I40836
R:Reinscheid, D.J.; Eikmanns, B.J.; Sahm, H.
Microbiology 140, 3099-3108, 1994
A:Title: Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene
A:Reference number: I40715; MUID:95111631; PMID:7812449
A:Accession: I40715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RES>
A:Cross-references: EMBL:X78491; NID:g530011; PIDN:CAA55243.1; PID:g530012
R:Lee, H.
J. Microbiol. Biotechnol. 4, 256-263, 1994
A:Title: Molecular characterization of aceB, a gene encoding malate synthase in Corynebacterium
A:Reference number: I40836
A:Accession: I40836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-739 <RE2>
A:Cross-references: GB:I27123; NID:g853573; PIDN:AAA68074.1; PID:g853574

C;Genetics:
A:Gene: aceB

C;Genetics:
A;Gene: aceB
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 37.8%; Score 2289; DB 2; Length 739;
Best Local Similarity 61.3%; Pred. No. 1.5e-129;
Matches 444; Conservative 90; Mismatches 184; Indels

QY	9	TD	VS	GN	L	R	A	R	V	L	P	D	P	S	W	A	G	V	K	V	A	D	L	T	P	Q	N	A	L	L	N	A	R	D	E	L	Q	68
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	20	T	B	V	D	A	G	M	O	V	A	K	V	L	D	F	V	T	E	A	V	L	P	R	V	G	D	A	E	F	W	G	F	A	A	L	D	79
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	69	A	O	I	D	K	W	H	R	R	R	V	E	I	P	D	M	A	X	E	Q	L	T	E	G	I	L	L	P	P	D	O	F	T	I	T	S	128
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	80	M	L	I	D	Y	H	R	N	-	S	G	I	D	E	A	Y	S	D	F	L	K	E	L	G	I	V	L	E	P	E	A	E	I	R	T	138	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	129	L	N	A	R	P	A	L	N	A	A	R	M	G	S	L	D	A	L	T	G	D	V	I	P	E	T	D	G	A	E	K	T	P	N	K	V	188
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	139	L	N	A	R	P	A	L	N	A	A	R	M	G	S	L	D	A	L	T	G	N	A	I	P	E	T	D	G	A	E	K	K	E	I	N	P	198
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	189	L	S	S	G	F	G	D	A	T	E	T	V	O	D	G	O	L	V	A	L	P	K	S	T	G	L	A	N	C	P	A	G	E	T	G	A	248
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	199	L	D	G	A	S	H	A	D	E	K	Y	N	I	T	D	G	L	A	A	H	I	G	S	V	R	L	K	N	R	E	S	V	R	G	T	F	258
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	249	L	D	P	S	O	V	G	T	T	R	A	G	V	K	D	V	I	E	S	A	L	T	I	M	P	E	S	V	A	A	D	A	A	K	V	L	308
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	259	Q	I	D	P	V	H	P	I	G	K	A	D	K	T	G	D	V	I	E	S	A	L	T	I	M	P	E	S	V	A	A	D	A	E	K	T	318
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	309	A	V	D	K	O	G	T	A	F	L	R	V	L	N	R	D	N	Y	T	A	P	G	G	G	O	T	L	P	C	S	L	M	F	V	R	368	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	319	E	N	S	K	N	G	R	I	F	T	R	E	L	N	K	D	R	V	I</																		

RESULT 8

malate synthase B2123 [imported] - Bacillus halodurans (strain C-125)
E83916
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83916
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A8350; MUID:20512592; PMID:11058132
A:Accession: E83916
A>Status: preliminary

A;Molecule type: DNA
A.Residues: 1-727

A:Molecule type: DNA
A:Residues: 1-727 <STO>
A:Cross-references: GB:AP001514; GB
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2133

Query Match 37.3%; Score 2258; DB 2; Length 727;
Best Local Similarity 59.9%; Pred. No. 1.1e-127;
Matches 441; Conservative 97; Mismatches 186; Indels 1

Qy	1	MOHHHHHTDRVSGVGNLRIARVLVDFVNNBALPGTDIDPDSFWAGVYDKVVDLTTPQOQAL	60
Db	1	METYEH-----VGNLOVATKLITFEQALPGTGIKKDDFWSGVQOQITILMPENKML	53
Qy	61	LNARDELQAQIDKWHRRRVEIPIDMDAYROFTEIGYLLPEPDFTITTSGVDAEITTTTA	120
Db	54	LAKREEIQATIDAMHQNK-CPIDFSAYHSFLEEIGVLEPIPEHVITITENVDDEIAAQA	112
Qy	121	GPOLVVPVLNARFALNANARWGLYDALYGTVDVITPETOGAEKGPYTNKVRGDKVIAYAR	180
Db	113	GPOLVVPVNNARYAINANARWGLYDALYGSNVISEEDCKETGYNPKRGTKVIOFAK	172
Qy	181	KFLDSDSVLSSGSGFDATGFTVQOQQLVWALPD-KSTGLANPCQFAGYTGAAESPSTSVLL	239
Db	173	DFLDHTFPPLTSGSHHEALNAYIMDKQLVVLSEKMKTRLKDETQFVYGQSGQDPSVILL	232
Qy	240	INHGHLHIIILDPSPQVQTTDRAGVKDVIILBSAITTIMDPEDSVAADAAKVLGVYRWML	299
Db	233	LHHGLFEVIQIDARHPIGKSDRAVKDILVLESALTIMDCEDSVAADAEKVVAVYQNWL	292
Qy	300	GLNKGDAAAADVDCGTAFRLVNRDRNVTAPGGQFTLPCRSLMFVRNVGHLMTNDIAVD	359
Db	293	GLMKGTLLEATPTKEGKTKKRLNEDRSYTAPNGETFSLPCRSLMFVRNVGHLMTTPVIRT	352
Qy	360	TTGSEVFPFGIMDALFTGLIIAHGLKASDVNGPLINRSTGSIYIVKPRMHGPAEVAFTCEL	419
Db	353	QSGSEVPEGILDGIVTSLIAKHDLIQ---NGTFRNSKKGSVYIVKPRMHGSEEVAFANEL	409
Qy	420	FSRVDVLGLPONTWKIGIMDEERTTNLKAICAKAADRWVFINGTFLDRTGDEHTISM	479
Db	410	FNRIEDILGLEBHTIKIGWDEERTSLNKAIEKVKERVVINTGTFLDRTGDEHTISM	469
Qy	480	EAGPMVRKGTWKMSQBWILAYEDHNVDAGLAGFSGRAQVGKGMWWTLMADMVETKIAQ	539
Db	470	EAGPMIRAGDKWSSSWLSAYERSNVAAGLTGCGFQGRAQIKGKNWAMPDMLNMEMEQKGTQ	529
Qy	540	PRAGASTAWPSPATATLHALHYHOVDVAAVQOGLAGKREATTEOLLTPIKLAKELAWAP	599
Db	530	LEAGANTAWPSPATATLHAHYHRHVPAIOKTLADDQTCYKEILEIPLAGSTDTWNE	589
Qy	600	BIREEVNNKOSILGYVYVVRVDQGVGSKVPDIHDVALMEDRATLRISSQLANWLHRGV	659
Db	590	DIQABELNNAOQILGYVYVVRVVEQIGGSKVPDIHNTALMEDRATLRISSSHANWLHRGI	649
Qy	660	ITSADVRLASLERMAPLNDRONAGVAVRPMAPFDSDIAFLAAQELILSGAQCPNGYTPP	719
Db	650	VSKEQVIRTEMKMAKVDEQAGDPATPRMADNLEQSAVAFQAALELVLKGTQEPSGITPP	709
Qy	720	ILHRRRRBPKARAEEK	735
Db	710	ILHRRRLBPKOKIAKE	725

RESULT 9

malate synthase (EC 4.1.3.2) isoenzyme G - *Escherichia coli* (strain K-12)
S51788
S51789
C:Species: *Escherichia coli*
C:Date: 23-Aug-1995 #sequence_revision 13-Mar-1997 #text_change 01-Mar-2002
C:Accession: S51788; PF5083
R:Molina, I.; Pellicer, M.T.; Badia, J.; Aguilar, J.; Baldona, L.
Eur. J. Biochem. 224: 541-548, 1994
A:Title: Molecular Characterization of *Escherichia coli* malate synthase G. Differentiation

A;Reference number: S51788; MUID:95010032; PMID:7925370

A;Accession: S51788

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-723 <MOL>

A;Cross-references: EMBL:X74547; NID:G517246; PIDN:CAAS2639.1; PID:G517247

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F65083

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-723 <BLAT>

A;Cross-references: GB:AE000380; GB:U00096; NID:G1789344; PIDN:AACT6012.1; PID:G1789348;

A;Title: The complete genome sequence of *Escherichia coli* K-12, substrain MGL655

C;Genetics:

A;Gene: glcB

C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 35.5%; Score 2154.5; DB 2; Length 723;

Best Local Similarity 58.0%; Pred. No. 1.8e-121;

Matches 419; Conservative 115; Mismatches 174; Indels 15; Gaps 5;

Qy 12 VSVGNRIARVLYDFVNNALPGTDIDPDPSFAGVDKVVADLTPOQALLNARDLQAI 71

Db 5 ITQSLRIRIDANFRFVDEEVLPGTGLDAAAFWRFDEIVHDLAPENRQLLAERDRIQAAL 64

Qy 72 DKHRRRVIEPIDMDAYRQFLTEIGVLLPEPDFTTTSGVDABITTTAGPOLVVPVLA 131

Db 65 DEHRSNPGPVKDAAYKSLRELGLVLPQPERVTEITGDSBITSGAPQLVWPANNA 124

Qy 132 RFALNANARWGLYDALYGVDPVETDGAEGKPTNKVGRDKVAYARKFDDSDVPLSS 191

Db 125 RYALNANARWGLYDALYGVDSIIPQGANVSG--YDPQGEQVIAWRFDESLPLEN 182

Qy 192 GSGGDATGTVQDGLVVALPD-KSTGLANPGFAGYTGAASTPSVLLINHLHIEILI 250

Db 183 GSYQDVAFVAVKQLRIKNGKETTTLTPAQFVGYRGAATAAPTCLIKNGLHIELQI 242

Qy 251 DPSOVTGTDRAQGVKDVILESAITTTMDREDSVAAVDAADKVLGVNMLKNGDLAAAV 310

Db 243 DANGRIKDDPAHINDVIEAALITLDCEDSVAAVDAEDKILLYKNLGLNGLQETLQENK 302

Qy 311 DKDGTAPLRVINDRYNTPAGGQFTLPGRSLMFVENVGHMTNDAIVDTDGESEVFEGIM 370

Db 303 EKNGRQIVRKINDRRHYTAADGSEISLHGRSLLFIRNVGHMTIPVIMWDSGNEIPEGIL 362

Qy 371 DALFTGLIAHGIKASDVNGPLINSGTGIYVKKPMHGPABVAFTCELFSRVEDVLGP 430

Db 363 DGWMTGAIALYDLKVK-----NSRTSGVYIVKPKMHPQEVAFANKLFTRIETMLGNA 416

Qy 431 QNTWKIGIMDEBERTTNLKAACIKAAADRVVFTINTGFLDRTGDEIHTSMEAGPMVRKGTM 490

Db 417 PNTLKWGIMDEBERTSILNRSCLQAARNRVAFINTGFLDRTGDMHVSWEAGPMLRKQM 476

Qy 491 KSPFWILAYEDHNVDAAGLAFSGRAQVKGKMWMTTELMAADVETKIAOPRAGASTAWVP 550

Db 477 KSTFWIKAYERNVLSGLFCGLRGKQAIKGKMWAMPDLNADMYSQKGDQLRAGANTAWVP 536

Qy 551 SPTAATLHALHYHOVDVAAVVOQGLA-----GKRATIEOLLTIPLAKELAWAPDIREEV 605

Db 537 SPTAATLHALHYHOTNVQSVQANIAQTEFWAFEPILLDDLLTIPVAENANNSAQIQEL 596

Qy 606 DNNQCSILGVYRWVQGVGCSKVPIHDVAMMEDRATLRISQLLANWLHGVITSADV 665

Db 597 DNNVQGLTGVVVRWVEQGIGCSKVPDIHNVAMMEDRATLRISQIHANWLHIGILTKEOV 656

Qy 666 RASLERMAPLVDRONAGDVAYRPMANFDDSTAFLLAAQELILSGAQQPNGYTEPILHR-R 724

Db 657 QASLENNAKVVDQONAGDPAYRPMAGNFANSCAFKAASDLIFLGKQPNGYTEPILLHAWR 716

Qy 725 RRE 727

Db 717 LRE 719

RESULT 10

F70794

probable ask protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C;Accession: F70794

R;Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70794

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <COL>

A;Cross-references: GB:AL022121; GB:AL123456; NID:G3261559; PIDN:CAA18031.1; PID:G2960133;

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: ask

C;Superfamily: aspartate kinase; aspartate kinase homology

F;3-406/Domain: aspartate kinase homology <DKI>

Query Match 26.7%; Score 1615.5; DB 2; Length 421;

Best Local Similarity 85.1%; Pred. No. 1.8e-89;

Matches 338; Conservative 7; Mismatches 35; Indels 17; Gaps 3;

Qy 749 VQYGGSSVADAEIRRVABEIRIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 808

Db 5 VQYGGSSVADAEIRRVABEIRIVATKKQNDVVVVSAMGDTTDDLLDLAQVCPAPPPR 64

Qy 809 ELDMLLAGRISNALVMAIESLGAHARSFTQAGVITGTHGNAKIIVTPGRLQTA 868

Db 65 ELDMLLAGRISNALVMAIESLGAHARSFTQAGVITGTHGNAKIIVTPGRLQTA 124

Qy 869 LEEGRVVLVAGFQGVSDTKDVTTLGGGSDTTAVAMAALGADVCEIYTDVIGIFSADP 928

Db 125 LEEGRVVLVAGFQGVSDTKDVTTLGGGSDTTAVAMAALGADVCEIYTDVIGIFSADP 184

Qy 929 RIVENARKLDTVTPEEMLEMAACGAKVLMURCEYARRHNPVHVRSSYSDRPTVVVGS 988

Db 185 RIVENARKLDTVTPEEMLEMAACGAKVLMURCEYARRHNPVHVRSSYSDRPTVVVGS 244

Qy 989 IKQVMEDEPILTVGAHRSSEAKVTIVGLPDIPGYAAKVFRAARRRQHRHGAEE----- 1043

Db 245 IKQVMEDEPILTVGAHRSSEAKVTIVGLPDIPGYAAKVFRAA-----DADVNIDM 295

Qy 1044 RLQGRGQDRHLLHLLPQTS---GPPPKWKTSETRSASTQLLYDDHIGKVSILIGAGMR 1100

Db 296 VLQNVSKVEDGKTDITFTCSRVDVGAPEKLSLNEIGFSQLLYDDHIGKVSILIGAGMR 355

Qy 1101 SHPQVATTFCEALAAVGVNIELISTSDQSRSCAAT 1137

Db 356 SHPQVATTFCEALAAVGVNIELISTSEIRISVILCRDT 392

RESULT 11

GB7199

aspartokinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 14-Dec-2001

C;Accession: GB7199

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: G87199
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:G13093941; PIDN:CAC31839.1; GSPDB:GN00147
C:Genetics:
A:Gene: ask
C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 25.5%; Score 1545.5; DB 2; Length 421;
Best Local Similarity 80.4%; Pred. No. 2.9e-85;
Matches 319; Conservative 18; Mismatches 43; Indels 17; Gaps 3;

QY 749 VQYGGSSVADAEIRIRVAERIVATKKQGNDDVVVVVVSANGDITDLDLLAQVCPAPP 808
DB 5 VQYGGSSVADAEIRIRVAERIVQTKQGNDDVVVVVVSANGDITDLDLLAQVCPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPEGLQTA 868
DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPEGLQTA 124

QY 869 LEEGRVVLVAGFGQVSDTKDVTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGRVVLVAGFGQVSDTKDVTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGIFSADP 184

QY 929 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 244

QY 989 IKDVMEDPILTGVAHDSREAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 1043
DB 245 IKDVMEDPILTGVAHDSREAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----DADVNIDM 295

QY 1044 RLQGRGRQDRHLLH---LPQTSQPPPKWNTSRSEASTQLLYDDHIGKVSILGAGWR 1100
DB 296 VLQNVSKVEDGKTDITFTCSRDSPGPIAAKGLSLRDEIGFTQLLYDDHIGKVSILGAGWR 355

QY 1101 SPFGVTATFCEALAAVGVNIELISTSESDQRSCCAAT 1137
DB 356 SPFGVTATFCEALAAVGVNIELISTSEIRISVLCDRT 392

RESULT 12

S42422
aspartate kinase (EC 2.7.2.4) alpha chain - Mycobacterium smegmatis
N:Alternate names: aspartokinase alpha chain
C:Species: Mycobacterium smegmatis
C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 18-Jun-1999
C:Accession: S42422; S31801
R:Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A:Title: Isolation and characterization of the aspartokinase and aspartate semialdehyde
F:1-421/Product: aspartate kinase alpha chain #status predicted <ASA>
A:Reference number: S42421; MUID:194254720; PMID:7910936
A:Accession: S42422
A:Molecule type: DNA
A:Residues: 1-421 <CIR>
A:Cross-references: EMBL:Z17372; NID:G44506; PIDN:CAA78984.1; PID:G581352
A:Note: the authors translated the initiation codon GTG for residue 1 as val
C:Genetics:
A:Start codon: GTG
C:Superfamily: aspartate kinase; aspartate kinase homology
C:Keywords: alternative initiators; phosphotransferase
F:1-421/Product: aspartate kinase alpha chain #status predicted <ASA>
F:3-406/Domain: aspartate kinase homology <DKI>
F:250-421/Product: aspartate kinase beta chain #status predicted <ASB>

Query Match 24.4%; Score 1478.5; DB 2; Length 421;
Best Local Similarity 77.6%; Pred. No. 3.1e-81;
Matches 302; Conservative 31; Mismatches 35; Indels 21; Gaps 4;

QY 749 VQYGGSSVADAEIRIRVAERIVATKKQGNDDVVVVVVSANGDITDLDLLAQVCPAPP 808

DB 5 VQYGGSSVADAEIRIRVAERIVETKAGNDVVVVVVSANGDITDLDLLAQVSPAPP 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPEGLQTA 868
DB 65 EMDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPEGLQTA 124

QY 869 LEEGRVVLVAGFGQVSDTKDVTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGIFSADP 928
DB 125 LDEGRVVLVAGFGQVSDTKDVTTLGRGSGDITAVAMAAALGADVCEIYTDVDFGIFSADP 184

QY 929 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVNARKLDTVTFFEMLEMAACGAKVLMRCVEYARRHNPVHVRSSYSDRPGTVVVG 244

QY 989 IKDVMEDPILTGVAHDSREAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE----- 1043
DB 245 IKDVMEDPILTGVAHDSREAKVTIVGLPDIPGYAAKVFRAVARRRQHRHGAEE-----ADVNDM 295

QY 1044 RLQGRGRQDRHLLH---LPQTSQPPPKWNTSRSEASTQLLYDDHIGKVSILGAG 1098
DB 296 VLQNVSKVEDGKTDITFTCSRDSPGPIAAKGLSLRDEIGFTQLLYDDHIGKVSILGAG 353

QY 1099 MRSPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 MRSPGVTATFCEALAAVGVNIELISTSE 382

RESULT 13

S15276
aspartate kinase (EC 2.7.2.4) alpha chain - Corynebacterium glutamicum
N:Alternate names: aspartokinase alpha chain
C:Species: Corynebacterium glutamicum
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Jun-1999
C:Accession: S15276; S49977
R:Kalinowski, J.; Cremer, J.; Bachmann, B.; Eggeling, L.; Sahm, H.; Puhler, A.
Mol. Microbiol. 5, 1197-1204, 1991
A:Title: Genetic and biochemical analysis of the aspartokinase from Corynebacterium glut
A:Reference number: S15276; MUID:92065816; PMID:19556296
A:Accession: S15276
A:Molecule type: DNA
A:Residues: 1-421 <PUE>
A:Cross-references: EMBL:X57226; NID:G40509; PIDN:CAA40502.1; PID:G580983
R:Experimental source: ATCC 13032
R:Serebrjanski, I.; Wojcik, F.; Reyes, O.; Leblon, G.
submitted to the EMBL Data Library, November 1994
A:Description: Two loci of Corynebacterium glutamicum ATCC17965 that complement Escheric
A:Reference number: S49977
A:Accession: S49977
A:Molecule type: DNA
A:Residues: 329-421 <SER>
A:Cross-references: EMBL:X82928; NID:G599716; PIDN:CAA58100.1; PID:G599717
R:Kalinowski, J.; Bachmann, B.; Thietbach, G.; Puhler, A.
Mol. Gen. Genet. 224, 317-324, 1990
A:Title: Aspartokinase genes lysC-alpha and lysC-beta overlap and are adjacent to the as
A:Reference number: S12250; MUID:91094767; PMID:1980002
A:Contents: annotation
C:Genetics:
A:Gene: lysC-alpha
A:Start codon: GTG
C:Superfamily: aspartate kinase; aspartate kinase homology
C:Keywords: alternative initiators; phosphotransferase
F:3-406/Domain: aspartate kinase homology <DKI>

Query Match 20.3%; Score 1229.5; DB 2; Length 421;
Best Local Similarity 64.3%; Pred. No. 2.7e-55;
Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

QY 749 VQYGGSSVADAEIRIRVAERIVATKKQGNDDVVVVVVSANGDITDLDLLAQVCPAPP 808
DB 5 VQYGGSSVADAEIRIRVAERIVATKKAGNDVVVVVVSANGDITDLDLELAANVPVPPAR 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTTGTHGNAKIIDVTPEGLQTA 868

Db 65 EMDMLTAGERISNALVMAIESLGAEQFTSQAGVLTTERHGNAIVDVTGVRREA 124
 Qy 869 LEGRVVLVAGFGVSDTQDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 928
 Db 125 LDEGKICIVAGFGVQNKETEDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGVTADP 184
 Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTIVVGS 988
 Db 185 RIVPNAOKLEKIFEEMLAAGVSKILVRSVEYARRHNPVHVRSYSDRPGTIVVGS 244
 Qy 989 IKDVPMDPILITGVADHRSSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Qy 1039 HGAARLQGRGROD-RHLLHLP--QTSGPPPPWKNWTRSETRASQTLIDYDHHGKVSU 1095
 Db 305 DGTDTITFTCFRSDGRAMEILKLVQVQ- - - - -TNVLYDDQVGVKSLV 350
 Qy 1096 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 382

RESULT 14
 T35383
 probable aspartokinase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 31-Jan-2000
 C:Accession: T35383
 R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z21576
 A:Accession: T35383
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-425 <MUR>
 A:Cross-references: EMBL:AL079348; PIDN:CAB45482.1; GSPDB:GN00070; SC06T3.26C
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: ask; SC06T3.26C
 C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 18.8%; Score 1141; DB 2; Length 425;
 Best Local Similarity 61.6%; Pred. No. 5.7e-61;
 Matches 239; Conservative 54; Mismatches 79; Indels 16; Gaps 4;

Qy 749 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
 Db 5 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
 Qy 809 ELDMLTAGERISNALVMAIESLGAEQFTSQAGVLTTERHGNAIVDVTGVRREA 124
 Db 65 ELDMLTAGERISNALVMAIESLGAEQFTSQAGVLTTERHGNAIVDVTGVRREA 124
 Qy 869 LEGRVVLVAGFGVSDTQDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 928
 Db 125 LDEGKICIVAGFGVQNKETEDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGVTADP 184
 Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTIVVGS 988
 Db 185 RIVPNAOKLEKIFEEMLAAGVSKILVRSVEYARRHNPVHVRSYSDRPGTIVVGS 244
 Qy 989 IKDVPMDPILITGVADHRSSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Qy 1039 HGAARLQGRGROD-RHLLHLP--QTSGPPPPWKNWTRSETRASQTLIDYDHHGKVSU 1095
 Db 305 DGTDTITFTCFRSDGRAMEILKLVQVQ- - - - -TNVLYDDQVGVKSLV 350
 Qy 1096 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 382

Search completed: November 21, 2003, 16:09:56
 Job time : 26.4499 secs

Db 358 KSNFGVTDFTALSDAGVNIELISTSE 385

RESULT 15
 C33531
 aspartate kinase alpha and beta chain PA0904 [imported] - Pseudomonas aeruginosa (strain
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 14-Dec-2001
 C:Accession: C33531
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br-
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: C33531
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-412 <STO>
 A:Cross-references: GB:AE004525; GB:AE004091; NID:g9946805; PIDN:AAG04293.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: lysC; PA0904
 C:Superfamily: aspartate kinase; aspartate kinase homology

Query Match 13.6%; Score 823; DB 2; Length 412;
 Best Local Similarity 42.6%; Pred. No. 6.8e-42;
 Matches 170; Conservative 80; Mismatches 105; Indels 44; Gaps 4;

Qy 749 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
 Db 5 VQYGGSSVADAEIRRVARRIVATKQGNVAVVVSAMGDTTDDLLDLAQVCPAPP 808
 Qy 809 ELDMLTAGERISNALVMAIESLGAEQFTSQAGVLTTERHGNAIVDVTGVRREA 124
 Db 65 ELDMLTAGERISNALVMAIESLGAEQFTSQAGVLTTERHGNAIVDVTGVRREA 124
 Qy 869 LEGRVVLVAGFGVSDTQDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGFSADP 928
 Db 125 LDEGKICIVAGFGVQNKETEDVTTLGRGSGDTTAVAMAAALGADVCEIYTDVDFGVTADP 184
 Qy 929 RIVNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNPVHVRSYSDRPGTIVVGS 988
 Db 185 RIVPNAOKLEKIFEEMLAAGVSKILVRSVEYARRHNPVHVRSYSDRPGTIVVGS 244
 Qy 989 IKDVPMDPILITGVADHRSSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Db 245 MEDIPVEEAVLTGATDKSEAKVTIVGLPIGVAARVFRVAVRA- - - - -RRRQRH 1038
 Qy 1039 HGAARLQGRGROD-RHLLHLP--QTSGPPPPWKNWTRSETRASQTLIDYDHHGKVSU 1095
 Db 305 DGTDTITFTCFRSDGRAMEILKLVQVQ- - - - -TNVLYDDQVGVKSLV 350
 Qy 1096 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 1127
 Db 351 GAGVRSHPGVTATFCEALAAVGVNIELISTSE 382

Search completed: November 21, 2003, 16:09:56
 Job time : 26.4499 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:51:11 ; Search time 13.8195 Seconds
(without alignments)
3988.226 Million cell updates/sec

Title: US-09-688-672A-52

Perfect score: 6061
Sequence: 1 MQHHHHHTDRSVGNURIA.....SAATRRPCTRGDRGWACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3810	62.9	741	1	MASZ_MYCTU
2	3117.5	51.4	731	1	MASZ_MYCLE
3	2634	43.5	724	1	MASZ_RHOFA
4	2569	42.4	725	1	MASZ_PSEAE
5	2541	41.9	725	1	MASZ_PSEPK
6	2497	41.2	725	1	MASZ_PSEEL
7	2489	41.1	725	1	MAZI_PSESM
8	2416	39.9	721	1	MASZ_RHILO
9	2380	39.3	731	1	MASZ_AGRTS
10	2367.5	39.1	723	1	MASZ_RHILV
11	2351.5	38.8	723	1	MASZ_RHIME
12	2318	38.2	728	1	MASZ_BRUME
13	2313	38.2	728	1	MASZ_BRUSU
14	2289	37.8	738	1	MASZ_COREL
15	2273	37.5	741	1	MASZ_COREF
16	2272.5	37.5	721	1	MASZ_BRAVA
17	2258	37.3	727	1	MASZ_BACHD
18	2215	36.5	725	1	MAZ2_PSESM
19	2168.5	35.8	722	1	MASZ_ECOL6
20	2156.5	35.6	722	1	MASZ_SHIPL
21	2154.5	35.5	722	1	MASZ_ECOLI
22	1615.5	26.7	421	1	AK MYCTU
23	1478.5	24.4	421	1	AK MYCSM
24	1248.5	20.6	421	1	AK COREF
25	1224.5	20.2	421	1	AK CORGL
26	1216.5	20.1	421	1	AK CORFL
27	823	13.6	412	1	AK_PSEAE
28	819	13.5	405	1	AK_TRETH
29	785	13.0	411	1	AK2_BACST
30	778	12.8	407	1	AK2_BACSP
31	734.5	12.1	415	1	AK_AOUAE
32	721	11.9	405	1	AK_HELPJ
33	712.5	11.8	408	1	AK2_BACSU

34	711	11.7	405	1	AK_HELPY
35	532	8.8	404	1	AK1_BACSU
36	492	8.1	473	1	AK_VETJA
37	480.5	7.9	446	1	AK_RICPR
38	369.5	6.1	819	1	AKIH_SERMA
39	364.5	6.0	820	1	AKIH_ECOLI
40	324	5.3	921	1	AKH_DAUCA
41	319.5	5.3	815	1	AKH_HABIN
42	310.5	5.1	449	1	AK3_ECOLI
43	308.5	5.1	454	1	AK3_BACSU
44	285	4.7	816	1	AKH_BUCAI
45	282	4.7	431	1	AK_CHLTR

ALIGNMENTS

RESULT 1

ID	MASZ_MYCTU	STANDARD;	PRT;	741 AA.
AC	Q50596;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Malate synthase G (EC 2.3.3.9).			
GN	GLCB OR RV1837C OR MT1885 OR MTCY1A11.06.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaseae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Dalcher A., Uutterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +			
CC	CoA.			
CC	-1- PATHWAY: Glyoxylate bypass; second step.			
CC	-1- SUBUNIT: Monomer (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to the malate synthase family. GlsC subfamily.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z78020; CAB01465.1; -			
DR	EMBL; AE007047; AAK46156.1; -			

```

DR PIR; F70722; F70722.
DR PDB; 1N81; 18-DEC-02.
DR PDB; 1N8W; 18-DEC-02.
DR TIGR; MT1885; -.
DR Tuberculist; Rv1837c; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn G; 1.
DR TIGRFAMs; TIGR01345; malate_syn G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome; 3D-structure.
FT ACT_SITE 339 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 633 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 741 AA; 80403 MW; A92F54E0FE9B7C64 CRC64;

Query Match 62.9%; Score 3810; DB 1; Length 741;
Best Local Similarity 100.0%; Pred. No. 1.5e-205;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TDRVSVGNRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 2 TDRVSVGNRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSGVDAEITTTAGPQLVVPV 128
DB 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSGVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANARWGLYDALYGTVDIPETDGAEGPTNKVRGKVAYARFLLDDSV 188
DB 122 LNARFALNANARWGLYDALYGTVDIPETDGAEGPTNKVRGKVAYARFLLDDSV 181
QY 189 LSSGSGFGDATGFTVQDQQLVWALPKDSTGLANGPQAGTGAESPTSVLLINHLGHEI 248
DB 182 LSSGSGFGDATGFTVQDQQLVWALPKDSTGLANGPQAGTGAESPTSVLLINHLGHEI 241
QY 249 LIDPESQVGTTRAGVKVDVILSAITTIMDFEDSVAADVAADKVLGYRNWGLNGKDLAA 308
DB 242 LIDPESQVGTTRAGVKVDVILSAITTIMDFEDSVAADVAADKVLGYRNWGLNGKDLAA 301
QY 309 AVDKDGTAFILVLRNDRNTAPGGGFTLPGSLMFVRNVGHLMNDATVDTDSSEVFEG 368
DB 302 AVDKDGTAFILVLRNDRNTAPGGGFTLPGSLMFVRNVGHLMNDATVDTDSSEVFEG 361
QY 369 IMDALFTGLIAIHLKASDVNGPLINSGTSYIVKPKMHPAEVAFTCELSRVEDVLG 428
DB 362 IMDALFTGLIAIHLKASDVNGPLINSGTSYIVKPKMHPAEVAFTCELSRVEDVLG 421
QY 429 LPONTMKIGIMDEERTTNLAKIAAADRVVINTGFLDRTGDEIHTSMAGPMVRKG 488
DB 422 LPONTMKIGIMDEERTTNLAKIAAADRVVINTGFLDRTGDEIHTSMAGPMVRKG 481
QY 489 TMKSQWILAYEDHNVDAGLAAGFSGRAQVKGKMTMTLMADWVETKIAQPRAGASTAW 548
DB 482 TMKSQWILAYEDHNVDAGLAAGFSGRAQVKGKMTMTLMADWVETKIAQPRAGASTAW 541
QY 549 VPSPTAATHALHYQVDVAAVQOGLAGKRRTATIEQLTIPIAKELANAPDSIREVNN 608
DB 542 VPSPTAATHALHYQVDVAAVQOGLAGKRRTATIEQLTIPIAKELANAPDSIREVNN 601
QY 609 CQSILGVVVRWDQVGCCKVPDIHVDALMEDRATLRISSQLLANLWLRHGVITSADVRAS 668
DB 602 CQSILGVVVRWDQVGCCKVPDIHVDALMEDRATLRISSQLLANLWLRHGVITSADVRAS 661
QY 669 LERMAPLVDRQAGDVAYRPMAPNFDDSTAFIAQELILSGAQQNGVTEPILHRRREF 728
DB 662 LERMAPLVDRQAGDVAYRPMAPNFDDSTAFIAQELILSGAQQNGVTEPILHRRREF 721
QY 729 KARAEEKPAPSDRAGDAAAR 748
DB 722 KARAEEKPAPSDRAGDAAAR 741

```

```

RESULT 2
MASZ MYCLE
ID MASZ MYCLE STANDARD; PRT; 731 AA.
AC O32913;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR ML2069 OR MLCBL788.27.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hanlin N.,
RA Holroyd S., Hornsby I., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Batrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlsB subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL008609; CAAL5459.1; -
CC EMBL; AL583924; CAC31024.1; -
CC PIR; T44752; T44752.
CC HSP; P37330; 1D8C.
CC Leprosa; ML2069; -.
CC HAMAP; MF_00641; -.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthase.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_syn G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 638 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 731 AA; 80141 MW; 3878CADA45DB416C CRC64;

Query Match 51.4%; Score 3117.5; DB 1; Length 731;
Best Local Similarity 82.0%; Pred. No. 7.4e-167;
Matches 597; Conservative 54; Mismatches 72; Indels 5; Gaps 2;

QY 9 TDRVSVGNRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
DB 2 TDRVSVGNRIARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 61
QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSGVDAEITTTAGPQLVVPV 128
DB 62 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSGVDAEITTTAGPQLVVPV 121
QY 129 LNARFALNANARWGLYDALYGTVDIPETDGAEGPTNKVRGKVAYARFLLDDSV 188

```


Db 122 LNARFALNANRWGSLYDALYGTDTIPETEGAEGSEYKNGKIRGDKVIAVARKFMDQAVP 181
Qy 139 LSSGSGDATGFTVQDQGLVVALPDKTGLANPCQAGYGAASPT-SVLLINHLGHIE 247
Db 182 LASDWTNATGVSIFDQGLQIAITNSTGLASPEKFGYRNQLRSSNWSVLLANHLGHIE 241
Qy 248 ILIDPESOGVTTDRAGVKDVILESAITIMDFEDSVAADAAADKVLGRNWLKNGDGLA 307
Db 242 VLIDPESPIGKTPVGIKDVILESAITIMDFEDSVTAVDADKVGYNWLKNGDGLT 301
Qy 308 AAVDKDGTAFVLNDRNYTAPGGGFTLPGRSLMFRVNVGHUWMTDAIV---DPTGS 363
Db 302 EBNKDKGTFTRVLNADRSTYTPDGGELTLPGRSLLFVRNVGHUWMTDAIV---DPTGS 361
Qy 364 EVFEGIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMHPAEVAFTCELSRV 423
Db 362 EVFEGIDAVFTGLAHLGLTGTANGPLNSRTGSIYIVKPKMHPAEVAFTCELSRV 421
Qy 424 EDVLGLPONTMKGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGP 483
Db 422 EDVLGLPONTMKGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGP 481
Qy 484 MVRGTMKSQWILLAYEDHNVDAGLAAGFSGAQVGKGMWMTMELMADWVETKIAQPRAG 543
Db 482 MIRGAMKNSWIKAYEDANVDIGLAAGFGKGAQIGKGMWMTMELMADWVETKIAQPRAG 541
Qy 544 ASTAVPSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLTIPLAKELANAPDEIRE 603
Db 542 ATTAVPSPPTAATLHAMHYQVDVAAVQOGLTGORRATVDQLTIPLAKLANAPEIRE 601
Qy 604 EVDNQCQILGVYVRWVQGVGCKSPDIDHVALMEDRATURISSQLLANWLRHGVITSA 663
Db 602 EVDNQCQILGVYVRWVQGVGCKSPDIDHVALMEDRATURISSQLLANWLRHGVITSE 661
Qy 664 DVRASLERMAPLVDONAGDVAYRPMAPNFDDSTAFIAAQLILSGAQQNGVTEPIIHR 723
Db 662 DVRASLERMAPLVDONAGDVAYRPMAPNFDDSTAFIAAQLILSGAQQNGVTEPIIHR 721
Qy 724 RRRBFKAR 731
Db 722 RRRBFKAQ 729

RESULT 3
MASZ_RHOFA
ID MASZ_RHOFA STANDARD; PRT; 724 AA.
AC Q9AE55;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR VICA.
OS Rhodococcus fascians.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1828;
RN [1]_TaxID=1828;
RP SEQUENCE FROM N.A.
RC STRAIN=D188;
RA Verecke D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
RA Goethals K.;
RT "Characterization of a chromosomal locus that affects pathogenicity in
Rhodococcus fascians";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ301559; CAC35701.1; -
DR HSSP; P37330; 1D8C.
DR HAMAP; MF_00641; 1.
DR InterPro; IPR001465; Malate_synthase.
DR InterPro; IPR006253; Malate_synthG.
DR Pfam; PF01274; Malate_synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 724 AA; 78609 MW; F899FE88390995E CRC64;

Query Match 43.5%; Score 2634; DB 1; Length 724;
Best Local Similarity 69.0%; Pred. No. 7.5e-140;
Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;

Qy 9 TRVSVGNLRIRARVLYDFVNNALPGTDIDPDSFWAGVDKVVADLTQPONQALLNARDEIQ 68
Db 2 TRVQAGGQLQVAKVLFDFVEKEALPGTDIDPDSFAWAGAAVADLAPKXKALLAVRDEIQ 61
Qy 69 AQIDKMRHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDAAEITTTAGPQLVWPV 128
Db 62 GKVDAMHGEHAGAYDRAAYKAFLEIGYLLDEPADFQIHTSGVDTEITTTAGPQLVWPV 121
Qy 129 LNARFALNANRWGSLYDALYGTDTIPETDGAEGKPTNKVGRKVIAYARKFLDDSV 188
Db 122 LNARFALNANRWGSLYDALYGTDAIPETDGAEGKTSYKVRGDKVIAFARDFLDEALP 181
Qy 189 LSSGSGDATGFTVQDQGLVVALPDKST-GLANPGQFAGYTGAAESPTSVLLINHLGHIE 247
Db 182 LSSGSHVGTGVVDAASLTVLADGSTVGLKDPQLGQYGTPTDAPTALFVHNLHPE 241
Qy 248 ILIDPESOGVTTDRAGVKDVILESAITIMDFEDSVAADAAADKVLGRNWLKNGDGLA 307
Db 242 ILIDPESPIGKTDGAGVKDVILESAVTTIMDFEDSVAADADDKVLGRNWLKNGDGLT 301
Qy 308 AAVDKDGTAFVLNDRNYTAPGGGFTLPGRSLMFRVNVGHUWMTDAIVDSDGSEVE 367
Db 302 EBNKDKGTFTRAMNKDRTYISVDGSELTGHRSLLFVRNVGHUWMTDAIVDSDGSEVE 361
Qy 368 GIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMHPAEVAFTCELSRVEDVL 427
Db 362 GILDALFTSLAGLSLTPDNV---LSNRTGSLYIVKPKMHPDEVAFTAEELFGRVEQVL 418
Qy 428 GLPONTMKGIMDEERTTNLKAICAKAAADRVVFTINTGFLDRTGDEIHTSMEAGVNRK 487
Db 419 GLPNTLKVIMDEERTTNLKAICAAASERVVFTINTGFLDRTGDEIHTSMEAGVNRK 478
Qy 488 GTWKSOPWILLAYEDHNVDAGLAAGFSGAQVGKGMWMTMELMADWVETKIAQPRAGSTA 547
Db 479 GANKGKNTAAVEDFNVDTGLAGLQCKAQIGKGMWMTMELMADWVETKIAQPRAGSTA 538
Qy 548 WVPSPPTAATLHALHYQVDVAAVQOGLAGKRRATIEQLTIPLAKELANWLRHGVITSEVDN 607
Db 539 WVPSPPTAATLHALHYQVDVAFARQHEIAKAKRATVDEILEIPLAPSTDWTDDEKQNELDN 598
Qy 608 NCOSILGVYVRWVQGVGCKSPDIDHVALMEDRATURISSQLLANWLRHGVITSEVDN 667
Db 599 NSOSILGVYVRWVQGVGCKSPDIDHVALMEDRATURISSQFIANLWLRHGVITSEVDN 658
Qy 668 SLERMAPLVDONAGDVAYRPMAPNFDDSTAFIAAQLILSGAQQNGVTEPIIHRRRRE 727
Db 659 SLERMAPLVDONAGDVAYRPMAPNFDDSTAFIAAQLILSGAQQNGVTEPIIHRRRRE 718
Qy 728 PKA 730
Db 719 YKA 721

```

RESULT 4
MASZ_PSEAE STANDARD; PRT; 725 AA.
ID MASZ_PSEAE
AC Q91636;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR PA0482.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10994043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino B., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Felger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the malate synthase family. GLCB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB004485; AAC03871.1; -
DR PIR; H83586; H83586.
DR HAMAP; MF 00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRams; TIGR01345; Malate syn G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;
Query Match 42.4%; Score 2569; DB 1; Length 725;
Best Local Similarity 69.4%; Pred. No. 3.2e-136;
Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
QY 9 TDRVGNLRIARVLDFVNNALPGTDIDPDSFWAGVDKRVADLTPQNCALLNARDELQ 68
DB 2 TERVQVGLQVAKVLFDFVNNALPGTVGSADFTWGAENVNDLAPKNKALLAKREDELQ 61
QY 69 AQDKKRRRVEIPNDVAYRQLTEIGYLLPEPDDFTITTSVDAEITTTAGPOLVVPY 128
DB 62 AKDGHQARAGAHDAVAKYAFLEEGYLLPEADPQAGTQNVDDIARMAPGOLWVPV 121
QY 129 LNARFALNANRWGSLYDALYGTVDVTPDGAEGKPTNKVRGDKRVAYAKFLDDSDVP 188
DB 122 MNARFALNANRWGSLYDALYGTVDVISEGGAEKKGKNGKVRGDKRVAFARAFLEAP 181

```

```

189 LSSGFGDGTGFTYDQGLVVALPKS - TGLANPGFAGYTGAESPTSVLLINHLHIE 247
182 LESGSHVDATSYSVKNGALVVALKNGSETGLKNGAQFLAQGDAARQAVLLKHGLHFE 241
248 ILIDPESQVGTTRAGVKVIVLESATITIMDFSDSVAADVDAADKVLGYRWLGLNGDLA 307
242 IQIDFSSPVGQTDAAAGKVDVLMEEAALTTIMDCSDSVAADVDAADKVIYRWLGLMKGDLA 301
308 AAVDKDGTAFILVLRNDRNTAPGGGQFTLPGSLMFRVNVGHLMNDALVDTDGSEVPE 367
302 EVSKGGSTTRTNWPNRVYTRADGSLILHGLSLIFVRNVGHLMNDALDXDGEVPE 361
368 GIMDALFTGLIAIHGLKASDVNGPLI--NSRTGSIYIVKPMHGAFAVFTCLFSERVED 425
362 GIQDGLFTSLIAIH-----DLNGNTRSKNSRTGSVIVKPMHGPBEAAFTNLFQERVED 416
426 VLGLPONTMKIGIMDEERRITVNLKACIKAAADRVVFIINTGFLDRTGDEIHTSMEAGPMV 485
417 VLGLPNTLVKGMIMDEERRITVNLKACIKAAKORVVFIINTGFLDRTGDEIHTSMEAGAVV 476
486 RKGTKSQPWILAYEDHNDVDAAGFSGRAQVKGKMTMTTELMADMVETKIAQPRAGAS 545
477 RKGAMKSEKIVGAYENNVVDVGLATG-LQKRAQIGKGMWMPDLMAAMLEQKIGHPLAGAN 536
546 TAWVPSPTAATLHALHYHQVDVAQVQGLAGKRRATIEQLTTIPLAKELAWADETREV 605
537 TAWVPSPTAATLHALHYHKVDVFAQAEALAKRTPASVDDILTITPLAPNTNWTAEIKNEV 596
606 DNNQCSILGVYRWVDQGVCSKVPDIHDVVALMEDRATLRISQLLANLWRHGVITSADV 665
597 DNAQGILGVYRWVDQGVCSKVPDINDVGLMEDRATLRISQLLANLWRHGVISQEQV 656
666 RASLERMAPLVDRONAGDVAYRMAPNFDSDSIAPLAQELISLGAQPNQYTFPIHRRR 725
657 VESLKRMAVVVDRONADSPSYRMAPNFDNDVAFQALELVVEGTQPNQYTFPIHRRR 716
726 REFKA 731
717 REFKA 722
RESULT 5
MASZ_PSEPK STANDARD; PRT; 725 AA.
ID MASZ_PSEPK
AC Q8QX8;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR PP0356.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=150488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Seanan M., DeBoy R.T., Daugherty S., Kolchay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapfle E., Scanlan D., Tran K., Moazzez A.,
RA Uterback T., Rizzo M., Lee K., Kossack D., Moestl D., Wedler H.,
RA Lauber J., Sjepeandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eissen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer (By similarity).

```


Db 242 IQIDASTPYGQTDAGVKDVLMEALTTIMDCEDSVAAVDADKVKVIVYRNWGLMKGDIA 301
QY 308 AAVDKDGTAFRLVLRNDRNTYAPGGQFTLPGRSLMFVRNVGHLMNDIAVTDGSEVPE 367
Db 302 EEVAKGGKFTRTMNPDRVYTVGQDVTLHGKSLLFVRNVGHLMNDIAVTDGSEVPE 361
QY 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKWHGPAEVAFTCELFSEVEDVL 427
Db 362 GILDGGLTSLAAIHNLNGNTRS---NSRSGSYIVKPKWHGPAEVAFTCELFSEVEDVL 418
QY 428 GLPQNTMKIGIMDEERTTNLAKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGPMVRK 487
Db 419 NLPRNLTKVIGIMDEERTTNLAKACIKAAASERVFNITGFLDRTGDEIHTSMEAGPMVRK 478
QY 488 GTKSPWILAYEDHNVDAGLAFSGRAQVKGKMTTELADWVETKIAOPRAGASTA 547
Db 479 AAMKTEKWIAGYENWVNDIGLSTGLGQRAQIGKGMWAMPDLMAALEQKIAHPLAGANTA 538
QY 548 WVPSPATAATHALHYHVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIREVEDN 607
Db 539 WVPSPATAATHALHYHVDVFAQRAELAKERASVDDILTIPLAKNTOWSEIEINELDN 598
QY 608 NCOSILGYVVRVWDQGVGSKVPDIHDVALMEDRATLRISOLLANWLPHGVITSADVRA 667
Db 599 NAQGIILGYVVRVWDQGVGSKVPDINDVGLMEDRATLRISQHIANWLPHGVITQVLE 658
QY 668 SLERMAPLVDQRNAGVAVRPMAPNFDSDIAFLAQELISGAQPNGYTEPILHRRRRE 727
Db 659 SLKRMAPVVDQRNAGDALVRPLAPDFSDINIAQAAVELVIEGTQPNGYTEPILHRRRRE 718
QY 728 FKAR 731
Db 719 FKAK 722

RESULT 7

MAZI_PSESM STANDARD; PRT; 725 AA.
AC Q88AB2;
ID MASI_PSESM
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G 1 (EC 2.3.3.9).
GN GUCB1 OR GLCB-1 OR PSPT00480.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OX Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RI "Complete sequence of Pseudomonas syringae";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -|- PATHWAY: Glyoxylate bypass; second step.
CC -|- SUBUNIT: Monomer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the malate synthase family. GicB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMEL; AE016857; AAO54024.1; -.
DR TIGR; PSPT00480; -.
DR HAMAP; MF_00641; -. 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 340 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 631 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 725 AA; 79143 MW; F1993264E8083660 CRC64;
Query Match 41.1%; Score 2489; DB 1; Length 725;
Best Local Similarity 65.3%; Pred. No. 9.6e-132;
Matches 473; Conservative 97; Mismatches 150; Indels 4; Gaps 2;
QY 9 TDRVSVGNLRIARVLDVFNNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
Db 2 TEYVQVGDQLVARVLPDFVQNEATPGTVDAGAFWAGADOLIHDLAPKPKALLAQADELQ 61
QY 69 AQIDKWHRRVPIPDMDAYRQFLTEIGYLLPDPDDFTITTSVDABEITTTAGPQLVVPV 128
Db 62 AQIDAWHQRAGQAHDAAYKAFLEIGYLLPEAADFQITTONVDEIATMAGPQLVVPV 121
QY 129 LNAFPALNANRAGSLYDALYGTDPVETDGAEGKGTYNKVGDKVIAVARFELDSVP 188
Db 122 MNARFALNANRAGSLYDALYGTDAISEGGAEKGGYKVGDKVIAVARFELDSVP 181
QY 189 LSSGSGFDATGFTVQDQQLVVALPDKS-TGLANPGQFAGYTGAESPTSLLINHLHIE 247
Db 182 LAAGSHVDSTAYKLDIGRLVISLKGSGNTGLRDAQLVGFQGDASAPFAVLPHKNGLHPE 241
QY 248 ILIDPESQVGTDRAGVKDVILESATTTIMDFDSVAANDAAKVLGYRNWGLNKGDIA 307
Db 242 LQIDAAFPVQGTDPAGVKDIWMEAAUPTTINDCEDSIAAVDADDKVVVYRNWGLMKGDIV 301
QY 308 AAVDKDGTAFRLVLRNDRNTYAPGGQFTLPGRSLMFVRNVGHLMNDIAVTDGSEVPE 367
Db 302 EESVSKGGETRTMNPDRVYTVGQDVTLHGKSLLFVRNVGHLMNDIAVTDGSEVPE 361
QY 368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKWHGPAEVAFTCELFSEVEDVL 427
Db 362 GILDGGLTSLAAIHNLNGNTRS---NSRSGSYIVKPKWHGPAEVAFTCELFSEVEDVL 418
QY 428 GLPQNTMKIGIMDEERTTNLAKACIKAAADRVVFNITGFLDRTGDEIHTSMEAGPMVRK 487
Db 419 GLPNTLVKIGIMDEERTTNLAKACIKAAASERVFNITGFLDRTGDEIHTSMEAGPMVRK 478
QY 488 GTKSPWILAYEDHNVDAGLAFSGRAQVKGKMTTELADWVETKIAOPRAGASTA 547
Db 479 AAMKTEKWIAGYENWVNDIGLSTGLGQRAQIGKGMWAMPDLMAALEQKIAHPLAGANTA 538
QY 548 WVPSPATAATHALHYHVDVAAVQOGLAGKRRTIEQLLTIPLAKELAWAPDEIREVEDN 607
Db 539 WVPSPATAATHALHYHVDVFAQRAELAKERASVDDILTIPLAKNTOWSEIEINELDN 598
QY 608 NCOSILGYVVRVWDQGVGSKVPDIHDVALMEDRATLRISOLLANWLPHGVITSADVRA 667
Db 599 NAQGIILGYVVRVWDQGVGSKVPDINDVGLMEDRATLRISQHIANWLPHGVITQVLE 658
QY 668 SLERMAPLVDQRNAGVAVRPMAPNFDSDIAFLAQELISGAQPNGYTEPILHRRRRE 727
Db 659 SLKRMAPVVDQRNAGDALVRPLAPDFSDINIAQAAVELVIEGTQPNGYTEPILHRRRRE 718
QY 728 FKAR 731
Db 719 FKAK 722
RESULT 8
MAZI_RHILO STANDARD; PRT; 721 AA.
ID MASI_RHILO
AC Q98DK4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MLR4664.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OX Phyllobacteriaceae; Mesorhizobium.
OC NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Katanabe T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Katanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shampo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti".
RL DNA Res. 7:331-338(2000).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family, GLCB subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003004; BABS1267.1; -.
DR HAMAP; MF_00641; -; 1.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_synth_1.
KW Transferrase; Glyoxylate bypass; tricarboxylic acid cycle;
KW Complete proteome.
FT ACT SITE 338
FT ACT SITE 629
FT ACT SITE 629
SQ SEQUENCE 721 AA; 78058 MW; 65376311A7E1BDF CRC64;
Query Match 39.9%; Score 2416; DB 1; Length 721;
Best Local Similarity 64.4%; Fred. No. 1.1e-127;
Matches 467; Conservative 86; Mismatches 164; Indels 8; Gaps 3;
QY 9 TDRVSVGNLRIRARVLYDFVWNEALPGTDIDPDSFWAGVDKVVADLTPOQALLNARDELQ 68
Db 2 TDRIEIGLRIGAGELHDFVAGEALPGTGIAADATWAGFSAIVHDLAPKVRALLKKRAMQ 61
QY 69 AQIDKWHRRVIEPIDMAYRQFLTEIGYLPEDDDFTITTSQGVDAEITTTAGPOLVVPV 128
Db 62 ERLDQWYRDNG-APYDMEVYKFLKEIGYLPVEGFAFSVTDNDVPEIYVAGVQLVVPV 120
QY 129 LNARFALNAARWGSGLDLYLGTVDVIEPTGAEKGYPTNKGDKVIAYARKFLDSDVP 188
Db 121 MWARFALNAARWGSGLDLYLGTVDVIEPTGAEKKGFPNPAKVIAMAKOFLDOSVP 180
QY 189 LSSGFGDATGFTVDDGLVALPDKSTGLANPQFAGYTGAASTPSVLLINHLHIEI 248
Db 181 LTSQKWAGVGLSVAGHAKLKGAGAGGTTADPQFAGYRGDAANPDPAVLLVKNLHIEI 240
QY 249 LIDPSOYGTTRAGVKVILLESALTITMDPDSVAADADKVLGYRNWGLNKLGLAA 308
Db 241 VDRNNQIGRTDPAGIADVILLESALTITQDCDSVAADQKVVYRNWGLNKLGLAE 300
QY 309 AVDDKGTAFRLVNRDRNYTPAGGQFTLPGRSLMFVNRVGHMTNDAIVTDGSEVFEG 368
Db 301 BISKAGRSFVKLNADRYATAPAGQITVPGRSMLVNRVGHMTNPAIILDRDQNEVPEG 360

QY 369 IMDALFTGLIAHGLKASDV--NGPLINGRTGSIVYKPKMGHGPAAVFAFTCELFSEVEDV 426
Db 361 IMDAALTALIALH-----DVGPEGREANSRAGSMYVVKPKMGHGPAAVFAVEIFDSVEAL 415
QY 427 LGLPONTMKIGIMDEBERTTIVNLKACIKAAARVVFINTGFLDRTGDEIHTSMEAGPMVR 486
Db 416 LGVFRNTIKGIMDEBERTTIVNLKEAIRAAREVVFINTGFLDRTGDEIHTSMEAGPMVR 475
QY 487 KGTMSQPWILAYEDHNVDAGLAAGSGRAQVGKGMWMTMELMADVMVETKIAQPRAGAST 546
Db 476 KGMKQAAMISAYEANNVDTGLECGLAGHAQIGKGMWAMPDLMAMLEQKIAHPKAGANT 535
QY 547 AWVPSPTAATLHALHVQVDVAVQOGLAGKRRATIEQLTTIPLAKELAWADEIREVED 606
Db 536 AWVPSPTAATLHALHVQVDVAVQOGLAGKRRATIEQLTTIPLAKELAWADEIREVED 595
QY 607 NNCQSILGVYVRWVDGCGSKVPDIHDVVALMEDRATLRISOLLANLWLRHGVTTSADVR 666
Db 596 NNAQGLGVYVRWVDGCGSKVPDIHDVVALMEDRATLRISOLLANLWLRHGVTTSADVR 655
QY 667 ASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRR 726
Db 656 DSIQRMALVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRR 715
QY 727 EFXAR 731
Db 716 ELKAX 720
RESULT 9
MASZ_AGRTS STANDARD; PRT; 731 AA.
ID MASZ_AGRTS STANDARD; PRT; 731 AA.
AC Q8UJ85;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GCLCB OR ATU0047 OR AGR_C_78
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.P. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorllo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollman C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo J., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.

```
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008979; AAL41078.1; ALT INIT.
CC EMBL; AE007947; AAK85871.1; ALT_INIT.
CC PIR; AH2582; AH2582.
CC PIR; F97364; F97364.
CC HAMAP; MF_00641; -; 1.
CC InterPro; IPR001465; Malate_synthase.
CC InterPro; IPR006253; Malate_synthG.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_syn_G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
CC Complete proteome. 346
CC ACT_SITE 346 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 637 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 731 AA; 79520 MW; 69F304D5D6F8E8FFB CRC64;
Query Match 39.3%; Score 2380; DB 1; Length 731;
Best Local Similarity 63.5%; Pred. No. 1.2e-125;
Matches 461; Conservative 91; Mismatches 170; Indels 4; Gaps 2;
QY 8 HTRDSVGNLRARVLYDFVNNALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDEL 67
DB 9 HVSRTDKFGLSIDRLRYAFLTDEVLPGTGLDSETFEGPSAIVHSLSPKRELLAKRDAL 68
QY 68 QAQIDKWHRRVIEPDDMAYQFTEIGYLLAPEPDDFTITTSQVDARITTAGPOLVVP 127
DB 69 QEKIDGWYKENG-APGDFDAYFAFLKEIGYLLPEGGFGFVETNNVDPELVAVAGPOLVVP 127
QY 128 VLNAREPALNAANRWGSLYDALYDGVIPETDGAEGKPTYNKVRGDKVIAYARKEFLDVS 187
DB 128 VMNARYALNAANRWGSLYDALYDGVDAISDADGAEGKGRGNPKRGDKVIAYARNFLDESA 187
QY 188 PLSSGSGFATGFTVDDGQLVALPKSTGLANPGQFAGYTGAAESPTSVLLINGLHIE 247
DB 188 PLETGWSVDVTGFNIADGGLQLAIGAATGLDKDAVQFKGFGSEAAKPAATILGKNGLHTE 247
QY 248 ILIDPESQVGTDRAGVKDVIIESATITTMDFEDSVAADAAKVLGYRNMGLNKGDLA 307
DB 248 IVIDPSTEIGKSDRAGISDVILLESALTITTMDCEDSVAADKVLVYGNWGLNKGDLT 307
QY 308 AAVDKDGTAFRLVNLDRNYTAPGGQFTLPGRSLMFRVNVGHMTNDAIVDTDGSVEFE 367
DB 308 EAVSKGSGTFTRLNPDRTYTPDGSALTLPGRSLMFRVNVGHMTNPAILLDRDGRDVE 367
QY 368 GIMDALFTGLIAGHKASDVNGPLNSRTGISYIVKPKHGPAAEVAFTCELSFVEDVL 427
DB 368 GIMDAVTTALIALYDVGPS---GRRQNSRAGSVVYKPKHGPAAEVAFAFVAVENLV 424
QY 428 GLPQNTMKIGIMDEERTTWNLKACIKAAADRVVFTINTGFLDRTGDEIHTSMEAGPMVRK 487
DB 425 GNAPTNMGKIMDEERTTWNLKESITRAAKDRAVFTINTGFLDRTGDEIHTSMEAGPMIRK 484
QY 488 GTMKSQPIILAYEDHNVDAGLAAGSGRAOVKGKWWMTMELMADVETKIAQPRAGASTA 547
DB 485 GDNKQAAVIAAYENMNKVDVIGLECGSLGSHAQIKGKWWAMPDMLMALEQKIAHPKAGANTA 544
QY 548 WFPSPTAATLHALHYHQVDVAVQOGLAGKRATITTEQLTITPLAKELAWAPDEIRIEVDN 607
DB 545 WFPSPTAATLHATHYHKVDVAVQOGLKSGRAKLSDLISVFPVAPRPNWTPEIQRELDN 604
QY 608 NQSGILGYVVRWVDQGVGSKVPDHDVALMEDRATLRISQALLNWLHGHVITSDVRA 667
```

```
DB 605 NAOGILGYVVRWVDQGVGSKVPDINNIGLMDRATLRISAQHWANLHRGVVTEAQIIK 664
QY 668 SLERMAPLVRQVAGDVAVRPMAPNFDSDSTAFIAAQELILSGAQQNGVTEPILHRRRE 727
DB 665 TMKREMAVVDTONAGDFPAYLPMASDFDGSVAFOAAVELVLKGRQPNQNGVTEPVLHRRLE 724
QY 728 FKARAA 733
DB 725 LKAKQA 730
RESULT 10
MASZ RHILV
ID MASZ RHILV STANDARD; PRT; 723 AA.
AC Q937W7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR MASG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
CC NCBI_Taxid=387;
RN R01111;
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RT "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY059637; AAL17965.1; -
CC HAMAP; MF_00641; -; 1.
CC InterPro; IPR001465; Malate synthase.
CC InterPro; IPR006253; Malate_synthG.
CC Pfam; PF01274; Malate_synthase; 1.
CC TIGRFAMs; TIGR01345; malate_syn_G; 1.
CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
CC ACT_SITE 338 CATALYTIC BASE (BY SIMILARITY).
CC FT ACT_SITE 629 CATALYTIC ACID (BY SIMILARITY).
CC SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;
Query Match 39.1%; Score 2367.5; DB 1; Length 723;
Best Local Similarity 64.5%; Pred. No. 5.8e-125;
Matches 464; Conservative 92; Mismatches 173; Indels 5; Gaps 3;
QY 11 RVSYGNLRARVLYDFVNNALPGTDIDPDPSWAGVDKVVADLTPOQALLNARDELQAQ 70
DB 3 RVDKNGLALETVLHDFLVEVLPGLAVDADKFPDFSAIVHDLAPKNCALLAKRDELQVK 62
QY 71 IDKWHRRVIEPDDMAYQFTEIGYLLPEPDDFTITTSQVDARITTAGPOLVVPVLN 130
DB 63 IDWYRRHG-APADMDEYQSFLEIGYLLPEGSDFFQVSTQNVDPETASAGPOLVVPVWN 121
QY 131 ARPALNAANRWGSLYDALYDGVIPETDGAEGKPTYNKVRGDKVIAYARKEFLDSDVPLS 190
DB 122 ARVALNAANRWGSLYDALYDGVDAIPESDGAEGKSYNPKRGEKVIARVDFLDTAPLQ 181
QY 191 SGSGFDATGFTVDDGQLVALPKSTGLANPGQ- FAGYTGAESPTSVLLINHLGHIEIL 249
```

```

Db 182 DCRKWDGSAVKDGLVVRIDGEOAMLTGKHFAGYRGDAAPATHILLKNNGIHIEIV 241
Qy 250 IDPESQVOTDRAGVKDVLISAITTDMFDSVAADAAKVIKYRNWGLNKGDLAAA 309
Db 242 IDAATTIGKADSAHISDVWLESAITTDMCEDSAADDAEDKVVVYRNWGLNKGDLQEE 301
Qy 310 VDKDGTAFRLVLRNDRNTAPGGGFTLPGRSLMFVRNVGHMTNDATVDTDGSEVEPEGI 369
Db 302 VAKGTSIFRLNPLQVAGPDGAFAEVHRSRLMLVRNVGHMTNPAALLDRDGEVPEGI 361
Qy 370 MDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSFSDVVLGL 429
Db 362 MDAATIGLIALYDIPS---GRKXNSRTGSMVVKPKMHGPEEVAFAVEIFSRVEDALGL 418
Qy 430 PONTWKIGIMDEERTTNLAKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMVRKGT 489
Db 419 PRNTIKGIMDEERTTNLKECIARAERVRVFINTGFLDRTGDEIHTSMEAGPMVRKGD 478
Qy 490 MKSQPWIILAYEDHNVDAGLAAGFSGRAQVGKGMWMTMELMADVMETKIAQPRAGATAWV 549
Db 479 MRQAAMISAYENNVNVDIGLECLAGHAQIGKGMWMPDLMAAMLEQKIAHPKAGANTAWV 538
Qy 550 PSPTAATLHALHYHOVDVAANVOQGLAGKRRATITQLLTIPLAKELAVADPEIREVNNC 609
Db 539 PSPTAATLHALHYHRVNVARVQGLKDRARAKLSILSVPAVRPNWTPPEIQRELDNA 598
Qy 610 QSILGYVVRVVDQGVGCSKVPDIHDVALMEDRATLRISOLLANWLRHGVITSADVRASL 669
Db 599 QGILGYVVRVVDQGVGCSKVPDINNVLGMEBRATLRISAQMANWLRHGVITSAQIETW 659
Qy 670 ERMAPLVRONAGDVAYEPMAFNPDSDSIAFLAAQELILSGAQPNGYTEPILHRRRREFX 729
Db 659 RMAAVDRONASDPAYRPMAGNFSDSIAFOALDLVLKGRQPNGYTEPVLHRRRLELX 718
Qy 730 ABAA 733
Db 719 AKQA 722

RESULT 11
MASZ_RHIME
ID_MASZ_RHIME STANDARD; PRT; 723 AA.
AC Q92TA4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
EX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlsB subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
DR EMBL; AL591782; CAC41449.1; -.
DR HAMAP; MF_00641; -.
DR InterPro; IPR001465; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate syn G; 1.
KW Transferrase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 338 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 629 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 723 AA; 78853 MW; A0E95E8A5164B58 CRC64;
-----
Query Match 38.8%; Score 2351.5; DB 1; Length 723;
Best Local Similarity 62.9%; Pred. No. 4 6e-124;
Matches 455; Conservative 94; Mismatches 169; Indels 5; Gaps 3;

Qy 10 DRVSVGNLRIARVLYDFVNEALPGTDDIDPSFWAGVDKVVADLTPTQNALNARDELOA 69
Db 2 DRVEKYGLQIDAGLHRLFLVEEAMPGTGVDADRFFSAFSDLVHDLGPKNRALLVKRDELOA 61
Qy 70 QIDKHRRRRIEIPIDMDAYRQFLTEITGYLLPDPDDRTITTSVDAEITTTAGQOLVVPVL 129
Db 62 RLQDGTREHG-APVDMEAYEAFLEIREIGYLLPGPDYFVSTANVDSIATITAGQOLVVPVM 120
Qy 130 NARFALNAANARWGSLYDALYGTDIVPETDGAERKGTYNKVRGDKVIAYARFLLDSVPL 189
Db 121 NARYALNAANARWGSLYDALYGTDAIETDGAERKGTYNKVRGDKVIAYAREFLDASAPL 180
Qy 190 SSGSFGDATGFTVODQQLVVALPD-KSTGLANPGQFAGYTGAAESTSVLLNHGLHIEI 248
Db 181 AAGRWSDAKSFSEVAGATLVTLAGTKSAPRNSVQFAGYAGPAAPSEIVLRNGLHIVI 240
Qy 249 LIDPESQVGTTRAGVKDVLISAITTDMFDSVAADAAKVIKYRNWGLNKGDLAA 308
Db 241 VLDATTPIGKADAAGISDVVLESATITINDCEDSIAAADAEDKVIKYRNWGLNKGDLSE 300
Qy 309 AVDKDGTAFRLVLRNDRNTAPGGGFTLPGRSLMFVRNVGHMTNDATVDTDGSEVFE 368
Db 301 EVTKGGRAFTRRLNPDRAVTPADGATLTLFGRSLMLVRNVGHMTNPAVLDRDGEVPE 360
Qy 369 IMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMHGPAEVAFTCELSFSDVVLG 428
Db 361 LMDAMVTALIALHDIGR---NGRRANSRSGSMFVVRKPKMHGPEEVAFAVEIFARVAA 417
Qy 429 LPQNTWKIGIMDEERTTNLAKACIKAAADRVVFINTGFLDRTGDEIHTSMEAGPMVRK 488
Db 418 LPANAMKMGIMDEERTTNLKECIARAERVRVFINTGFLDRTGDEIHTSMEAGPMVRK 477
Qy 489 TMKSQPWIILAYEDHNVDAGLAAGFSGRAQVGKGMWMTMELMADVMETKIAQPRAGATA 548
Db 478 DMKQAPWISAYENNVNVDIGLECLAGHAQIGKGMWMPDLMAAMLEQKIAHPKAGANTAW 537
Qy 549 VPSPTAATLHALHYHOVDVAANVOQGLAGKRRATITQLLTIPLAKELAVADPEIREV 608
Db 538 VPSPTAATLHALHYHRVNVARVQGLKDRARAKLSILSVPAVRPNWTPPEIQRELDNN 597
Qy 609 QSILGYVVRVVDQGVGCSKVPDIHDVALMEDRATLRISOLLANWLRHGVITSADVRAS 668
Db 598 AQGILGYVVRVVDQGVGCSKVPDINNVLGMEBRATLRISAQMANWLRHGVITSAQIET 657
Qy 669 LESMAPLVRONAGDVAYRPMAGNFSDSIAFLAAQELILSGAQPNGYTEPILHRRRREF 728
Db 658 RMAAVDRONASDPAYRPMAGNFSDSIAFOALDLVLKGRQPNGYTEPVLHRRRLEL 717
Qy 729 KAR 731
-----

```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[illegible]

RESULT 14
MASZ_CORGL
ID_MASZ

```

Db 79 MLIDDDVHRNN-STIQEAYEDFKIGLVVEEFAAEIRTONQVDFEISTAGPOLWVEI 137
Qy 129 LNARFALNAANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 188
Db 138 LNARFALNAANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 197
Qy 189 LSSGSGDGTGFTVQDQGLVVALPKSTGLANPQAGTGAABSTPSVLLINHLGHIEI 248
Db 198 LGSASHADVEKNITDGLAAHIGDSYRLKNSRSGFTGNFLDPEAILLETNGLHIEL 257
Qy 249 LIDPESQVGTTRDRAGVKDVILESAITTIMDFEDSVAADAAKVLGYRNWGLNKGDLAA 308
Db 258 QIDPVPHIGKADTKGLDVLVESAITTIMDFEDSVAADAAKVLGYRNWGLNKGDLAA 317
Qy 309 AVDKDGTAFILVLRNDRNYTAPGGQFTLPGRSLMFRVNVGHLMNTDAIVDTGSEVFEG 368
Db 318 EMSKNGRIPTRELKNDKRVITGRNGTELVHLGRSLLFVRNVGHLMQNPSTL-IDGEEIPEG 376
Qy 369 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 428
Db 377 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 433
Qy 429 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKG 488
Db 434 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKA 493
Qy 489 TMSQOPWILAYEDHNVDAGLAAGFSRAQVKGKMTWMTLMADMVETKIAQPRAGASTAW 548
Db 494 DMQAPWKQAYEDNNVDAGIQRGLPKAQKGMWMTLMADMVETKIAQPRAGASTAW 553
Qy 549 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRATIEBOLLTIPLAKELAWAPDIREVDN 607
Db 554 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRATIEBOLLTIPLAKELAWAPDIREVDN 613
Qy 608 NCOSILGYVVRWVDQGVGSKVPDIHDVLMEDRATLRISQOLLANLWLRHGVITSADVRA 667
Db 614 NCOSILGYVVRWVDQGVGSKVPDIHDVLMEDRATLRISQOLLANLWLRHGVITSADVRA 673
Qy 668 SLERMAPLVDQRNAGDVAYEPMAFNPDSDTAFIAQELILSGAQOPNGYTEPILHRRRE 727
Db 674 SLERMAPLVDQRNAGDVAYEPMAFNPDSDTAFIAQELILSGAQOPNGYTEPILHRRRE 733
Qy 728 FKAR 731
Db 734 FKAK 737

```

RESULT 15
MASZ_COREF STANDARD; PRT; 748 AA.
AC Q8FNE3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GUCB OR MASZ OR CR2231.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriales; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.,
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CoA.
CC -!- PATHWAY: Glyoxylate bypass; second step.
CC -!- SUBUNIT: Monomer (By similarity).

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP005221; BAC19041.1; -
CC HAVAP; MF_008411; -; 1
CC InterPro: IPR001465; Malate synthase.
CC InterPro: IPR006253; Malate synthase.
CC Pfam: PF01274; Malate synthase; 1.
CC TIGRFAMs: TIGR01345; malate syn G; 1.
KW Transferase; Glyoxylate bypass; Tricarboxylic acid cycle;
KW Complete proteome.
FT ACT_SITE 362 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 653 CATALYTIC ACID (BY SIMILARITY).
SQ SEQUENCE 748 AA; 83491 MW; F9550473EC4E9A09 CRC64;
Query Match 37.5%; Score 2273; DB 1; Length 748;
Best Local Similarity 61.5%; Pred. No. 1.2e-119;
Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;
Qy 9 TDRVSVGNLRIARVLYDFVNEALPGTHIDPDSFWAGVKVADLTPOQALLNARDELQ 68
Db 26 TERYVGMQVAKVLRDLFTESVLPVRGVDAERFWNGFGDIVRDMTPRNLAREDELQ 85
Qy 69 AIDKWHRRVIEPDMDAYRQFLTEIGYLLPPDPFTITTSVDSEITTTAGPQLWVPV 128
Db 86 AQLDEYRENPGR-DPEKYEAFLEIGYLVDPAPAEIRTONIDSEIATTAGPQLWVPV 144
Qy 129 LNARFALNAANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 188
Db 145 LNARFALNAANARWGSYDALYGTVDIPETDGAEGKFTYKVRGDKVIAAKFDDSDVP 204
Qy 189 LSSGSGDGTGFTVQDQGLVVALPKSTGLANPQAGTGAABSTPSVLLINHLGHIEI 248
Db 205 LGSASHADVEKNITDGLAAHIGDSYRLKNSRSGFTGNFLDPEAILLETNGLHIEL 264
Qy 249 LIDPESQVGTTRDRAGVKDVILESAITTIMDFEDSVAADAAKVLGYRNWGLNKGDLAA 308
Db 265 QIDPVPHIGKADTKGLDVLVESAITTIMDFEDSVAADAAKVLGYRNWGLNKGDLAA 324
Qy 309 AVDKDGTAFILVLRNDRNYTAPGGQFTLPGRSLMFRVNVGHLMNTDAIVDTGSEVFEG 368
Db 325 EVAKGDRFTFRKNDKRVITGRNGTELVHLGRSLLFVRNVGHLMNTDAIVDTGSEVFEG 383
Qy 369 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 428
Db 384 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKMGPAEVAFTCELSRVEDVLG 440
Qy 429 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKG 488
Db 441 LPQNTMKIGIMDEERTTNVNLKACIKAAADRVVFINTEGFLDRTGDEIHTSMAGPMVRKA 500
Qy 489 TMSQOPWILAYEDHNVDAGLAAGFSRAQVKGKMTWMTLMADMVETKIAQPRAGASTAW 548
Db 501 DMQAPWKQAYEDNNVDAGIQRGLPKAQKGMWMTLMADMVETKIAQPRAGASTAW 560
Qy 549 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRATIEBOLLTIPLAKELAWAPDIREVDN 607
Db 561 VPSPTAATLHALHYQVDVAAVQOGL-AGKRRATIEBOLLTIPLAKELAWAPDIREVDN 620
Qy 608 NCOSILGYVVRWVDQGVGSKVPDIHDVLMEDRATLRISQOLLANLWLRHGVITSADVRA 667
Db 621 NCOSILGYVVRWVDQGVGSKVPDIHDVLMEDRATLRISQOLLANLWLRHGVITSADVRA 680
Qy 668 SLERMAPLVDQRNAGDVAYEPMAFNPDSDTAFIAQELILSGAQOPNGYTEPILHRRRE 727

```

Db 681 SLERMAVVVEQNAQDPNYLNNAPNFTESVAFQAARDLILKGTESPAGYTEPILHARRR 740

QY 728 FX 729

Db 741 FX 742

Search completed: November 21, 2003, 16:04:22
Job time : 15.8195 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 57.6698 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-52

Perfect score: 6061
Sequence: 1 MQHMHHTDRVSVGNLRIA.....SAATRRPCTGRDRMACQ 1172

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2634	43.5	724	2 Q9AE55	Q9ae55 rhodococcus
2	2569	42.4	725	16 Q9I636	Q9i636 pseudomonas
3	2416	39.9	721	16 Q98DK4	Q98dk4 rhizobium l
4	2380	39.3	744	16 Q9CJ85	Q9cjb5 agrobacteri
5	2367.5	39.1	723	2 Q937W7	Q937w7 rhizobium l
6	2351.5	38.8	723	16 Q92TA4	Q92ta4 rhizobium m
7	2318	38.2	728	16 Q9YIR3	Q9yir3 brucella me
8	2313	38.2	728	16 Q9FZ50	Q9fz50 brucella su
9	2273	37.5	748	16 Q9FN83	Q9fnb3 corynebacte
10	2258	37.3	727	16 Q9KB03	Q9kb03 bacillus ha
11	2168.5	35.8	723	16 Q9FDN6	Q9fdn6 escherichia
12	1545.5	25.5	421	16 Q9CB77	Q9cb77 mycobacteri
13	1272.5	21.0	421	2 Q8KQ29	Q8kq29 streptomyc
14	1267.5	20.9	421	2 Q9RQ25	Q9rq25 amycolatops
15	1232.5	20.3	421	2 Q93C57	Q93c57 corynebacte
16	1228.5	20.3	421	2 Q9AE85	Q9ae85 nocardia la

Q93C55 corynebacte
Q9xai7 streptomyc
Q8giq5 streptomyc
Q93tx4 stigmatella
P74569 synchocyst
Q8f565 leptospira
Q8ru02 methylobact
Q8dhw0 synchococc
Q8yrl4 anabaena sp
Q8gy66 bifidobacte
Q9jtn3 neisseria m
Q9jyn6 neisseria m
Q9kbb0 bacillus ha
Q9pnt4 campylobact
Q9wz17 thermotoga
Q9xik6 thermotoga
Q8y073 raistonia s
Q92ch4 listeria in
Q9a3w8 caulobacter
Q8y7n9 listeria mo
Q8y797 brucella me
Q8fyk1 brucella su
Q9ev19 bradyrhizob
Q98g91 rhizobium l
Q8usc0 agrobacteri
Q92mk3 rhizobium m
Q9kuh4 vibrio chol
Q8eb52 shewanella
Q8ra48 thermoanaer

ALIGNMENTS

RESULT 1

Q9AE55 ID Q9AE55 PRELIMINARY; PRT; 724 AA.
AC Q9AE55;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Malate synthase.
GN VICA.
OS Rhodococcus fascians.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1828;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D188;
RA Veresche D.M., Cornelis K., Van Montagu M., El Jaziri M., Holsters M.,
RA Goethals K.;
RT "Characterization of a chromosomal locus that affects pathogenicity in
RT Rhodococcus fascians.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301559; CAC35701.1; -;
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthG.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRFAMs; TIGR01345; malate_syn_G; 1.
SQ SEQUENCE 724 AA; 78609 MW; F889FE883890995E CRC64;
Query Match 43.5%; Score 2634; DB 2; Length 724;
Best Local Similarity 69.0%; Pred. No. 3.2e-144;
Matches 499; Conservative 82; Mismatches 138; Indels 4; Gaps 2;
QY 9 TDRVSVGNLRIRVLYDFVNEALPGDIDPDSFWAGVDKVVADLTQNALNARELQ 68
DB 2 TDRVQAGGLQVAKVLPDFVEKEALPGDLDSEAFWAGASVVIADLAEKALLAVRDEIQ 61
QY 69 AOIKDWRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVGVDATITTTAGPOLVVPV 128

Db 62 GKVDAMGEHAGAEYDRAAYKAFKEIGYLLDRPAPQIHTSGVDTEITTTAGPQLVVPV 121
 Qy 129 LNARFALNANARNGSLYDALYGTVDVTPETDGAEGKGYTNKVGDKVIAYARKFLDSDVP 188
 Db 122 LNARFALNANARNGSLYDALYGTDAIPETDGAEGKGYTNKVGDKVIAYARFDFLDEALP 181
 Qy 189 LSSGSFGDATGFTVQDQGLVVALPKST-GLANPGOPAGYTGAEBSPTSLLINHGHLIE 247
 Db 182 LSSGSHVGTGYVDAASLTVTLADGSTVGLKDFSQLGYQGTPDAPTLFFVHGLHFE 241
 Qy 248 ILIDPESQVGTDRAGVKDVIKESAITTMDFDSDVAADAKVILGYRNWLGKNGDLA 307
 Db 242 IQIDPESPIGKTGAGVKDVLLESATVTTIMDFDSDVAADAKVILGYRNWLGKNGDLT 301
 Qy 308 AAVDKDGTAFILVLRNDRNTAPGGQFTLPGSLMFRVNVGHMTNDALVDTDGSEVFE 367
 Db 302 EEVSKGKFTFRANKDKORTVTSVDGSELTLLHGRSLFVRNVGHMTNDALDADGNEVPE 361
 Qy 368 GIMDALFTGLIAHGLKASDVNGPLINSRTGSIYIVKPKMGHGAFAVFTCELFSSRVEDVL 427
 Db 362 GILDALFTSLAGLSLTPDNV---LNSRTGSLYIVKPKMGHGPDEVAFTAELEGRVEQVL 418
 Qy 428 GLPONTWKIGIMDEERTTNNLKACIKAAADRVVFTINTGLDRTGDEIHTSMEAGPMVRK 487
 Db 419 GLPTNTLKVIGIMDEERTTNNLKACIQAESRVVFTINTGLDRTGDEIHTSMEAGPMVRK 478
 Qy 488 GTMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADMTETKIAQPRAGASTA 547
 Db 479 GAMKGEKWIAYEDFNVDTGLAGLQKQAIKGMKAMPDLMLEQKIGHPKAGANTA 538
 Qy 548 WVPSPATAATLHALHYHGVDAVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREVDN 607
 Db 539 WVPSPATAATLHALHYHGVDAVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREVDN 607
 Qy 608 NCOSILGYVVRWIDQVGVCSKVPDIHDVVALMEDRATLRISSQLLANWLHGHVITSADVPA 667
 Db 599 NSQSIGLYVVRWIDQVGVCSKVPDIHDVVALMEDRATLRISSQLLANWLHGHVITSADVPA 667
 Qy 668 SLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 727
 Db 659 SLKRMVAVVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRRE 718
 Qy 728 FKA 730
 Db 719 YKA 721

RESULT 2

Q91636 PRELIMINARY; PRT; 725 AA.
 AC Q91636
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Malate synthase G.
 GN GLCB OR PA0482.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.B., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964 (2000).

DR EMBL; AE004485; AAG03871.1; --
 DR HSP; E37330; ID8C.
 DR InterPro; IPR001465; Malate synthase.
 DR InterPro; IPR006253; Malate synthase.
 DR Pfam; PF01274; Malate synthase; 1.
 DR TIGRfam; TIGR01345; malate_syn_G_1.
 KW Complete proteome.
 SQ SEQUENCE 725 AA; 78659 MW; 3669670A9E38D391 CRC64;
 Query Match 42.4%; Score 2569; DB 16; Length 725;
 Best Local Similarity 69.1%; Pred. No. 1.9e-140;
 Matches 502; Conservative 77; Mismatches 139; Indels 8; Gaps 3;
 Qy 9 TDRVSUGNLRTARVLYDFVNNELPGTIDDPDSFWAGVDKVVADLTQNCALLNARDELQ 68
 Db 2 TERVQVGLQVAKVLFDFVNNELPGTGVSDATFTWTGAEAVINDLAPKNKALLAKRDELQ 61
 Qy 69 AQIDKWHRRRRIEIDMDAYRQFLTEIGYLLPEPDDFTITTSVDADAITTTAGPQLVVPV 128
 Db 62 AKIDGWHQARAGQAHDVAAYKAFLEEIGYLLPEAEDFQAGTQNVDDDEIARMAGPQLVVPV 121
 Qy 129 LNARFALNANARNGSLYDALYGTVDVTPETDGAEGKGYTNKVGDKVIAYARKFLDSDVP 188
 Db 122 LNARFALNANARNGSLYDALYGTVDVTPETDGAEGKGYTNKVGDKVIAYARFDFLDEALP 181
 Qy 189 LSSGSFGDATGFTVQDQGLVVALPKS-TGLANPGOPAGYTGAEBSPTSLLINHGHLIE 247
 Db 182 LSSGSHVGTGYVDAASLTVTLADGSTVGLKDFSQLGYQGTPDAPTLFFVHGLHFE 241
 Qy 248 ILIDPESQVGTDRAGVKDVIKESAITTMDFDSDVAADAKVILGYRNWLGKNGDLA 307
 Db 242 IQIDPSSPVGQTDAGVKDVLMEALATITMDCEDSDVAADAKVILGYRNWLGKNGDLA 301
 Qy 308 AAVDKDGTAFILVLRNDRNTAPGGQFTLPGSLMFRVNVGHMTNDALVDTDGSEVFE 367
 Db 302 EEVSKGKFTFRANKDKORTVTSVDGSELTLLHGRSLFVRNVGHMTNDALDADGNEVPE 361
 Qy 368 GIMDALFTGLIAHGLKASDVNGPLI--NSRTGSIYIVKPKMGHGAFAVFTCELFSSRVED 425
 Db 362 GIQDGLFTSLIAH----DLNGNTSRKNGRTGSIYIVKPKMGHGAFAVFTCELFSSRVED 416
 Qy 426 VLGI.PONTWKIGIMDEERTTNNLKACIKAAADRVVFTINTGLDRTGDEIHTSMEAGPMV 485
 Db 417 VLGLPNTLKVIGIMDEERTTNNLKACIKAAADRVVFTINTGLDRTGDEIHTSMEAGPMV 476
 Qy 486 RGTMKSQPWILAYEDHNVDAGLAAGFSGRAQVKGMMTMTLMADMTETKIAQPRAGAS 545
 Db 477 RKGAMKSEKWIAYENNNDVGLATGLQKQAIKGMKAMPDLMLEQKIGHPKLAGAN 536
 Qy 546 TAWVPSPTAATLHALHYHGVDAVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREEV 605
 Db 537 TAWVPSPTAATLHALHYHGVDAVAAVQOGLAGKRRATIEQLTIPLAKELAWAPDEIREEV 605
 Qy 606 DNNCSILGYVVRWIDQVGVCSKVPDIHDVVALMEDRATLRISSQLLANWLHGHVITSADV 665
 Db 597 DNNCSILGYVVRWIDQVGVCSKVPDIHDVVALMEDRATLRISSQLLANWLHGHVITSADV 665
 Qy 666 RASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRR 725
 Db 657 VESLKRMAVVVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAQPNGYTEPILHRRR 716
 Qy 726 REFKA 731
 Db 717 REFKA 722
 RESULT 3
 Q98DK4
 ID Q98DK4 PRELIMINARY; PRT; 721 AA.
 AC Q98DK4
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)


```

141 VMARYALNANRWGSLYDALYGTDAISDADGAEKGRGNPKRGKVIAMARNFLDESA 200
188 PLSGSGFDATGTVDQGLVWALPKSGLANPGFAGYTGAAESPTSVLLINHLGHE 247
201 PLETGMSDVTGFNIADGLLQALGAATTLGLKDAVQFKGSGEAAKPAATILLGKGLHTE 260
248 ILDPESQVGTTRDRAGVKVILLESATTTMDPDSVAADKVLGYRNWGLKNGDLA 307
261 IVIDPSTEIKSRAGISDVILLESATTTMDCEDSVAAVDAEDKVLVGNWGLKRGDLT 320
308 AAVDKGTAFRLVNRDRNTYAPGGQFTLPGRSLMFRVNVGHMTNDIAVDTDSEVFE 367
321 EAVSKGCGNTFTRRLNPDYPTAPDGSALTLPGRSLMFRVNVGHMTNPAILDGRDVE 380
368 GIMDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGAFAVTCFLPSRVEDVL 427
381 GIMDAVVTAIALYDVGPS---GRRQNSRAGSIYIVKPKMGHGAFAVANEIFARVENV 437
428 GLPQNTMKIGIMDEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRK 487
438 GMAPNTMKGIMDEERTTIVNLKESIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMIRK 497
488 GTWKSQWILAYEDHNVDAAGLAFSGRAQVKGKMTMELMADVETKIAQPRAGASTA 547
498 GDMKQAIAIAYENWNVVDIGLECGLSHAQIGKGMWMPDLMAAMLEQKIAHPKAGANTA 557
548 WFSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIEEVDN 607
558 WFSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIEEVDN 617
608 NCQSILGYVVRWVDQVGCCKVDPIDHVALMEDRATLRISSQLLANWLRHGVITSADVRA 667
618 NAQSIGLYVVRWVDQVGCCKVDPIDHVALMEDRATLRISSQLLANWLRHGVITSADVRA 677
668 SLERMAPLVDRQVAGDVAYRPMAPNFDSDIARLAAQELILSGAQPNQGYTEPILHRRRE 727
678 TMKRAAVVDTQNGDPAYLPMSDFDGSVAFQAVELVLKRGEPNGYTEPVLHRRLE 737
728 FKAAA 733
738 LKAKQA 743

RESULT 5
Q937W7 PRELIMINARY; PRT; 723 AA.
ID AC Q937W7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Malate synthase G.
GN MASG.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF39;
RA Garcia de los Santos A., Hynes M.F.;
RL "Malate synthase gene from Rhizobium leguminosarum.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY059637; AAL17965.1; -
DR InterPro: IPR001465; Malate_synthase.
DR InterPro: IPR006253; Malate_synthg.
DR Pfam: PF01274; Malate_synthase; 1.
DR TIGRfams: TIGR01345; malate syn. G; 1.
SQ SEQUENCE 723 AA; 79677 MW; 4E879906CFD64444 CRC64;

Query Match 39.1%; Score 2367.5; DB 2; Length 723;
Best Local Similarity 64.1%; Pred. No. 9e-129;
Matches 464; Conservative 82; Mismatches 173; Indels 5; Gaps 3;

```

```

11 RVSGNLRIAEVLVDFVNNALPGTDIDPDSFWAGVDKVVADLTPOCALLNARDELQAO 70
3 RVDKNGLAETVLHDFLVEVLPGLAVDADKFPADPSAIVHDLAPKNCALLAKRDELQVK 62
71 IDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSVDABITTTAGPOLVVPVLN 130
63 IDWYRRHG-APADMDEYQSFLREIGYLLPEGSDFOVSTQVDPETIASIAGPOLVVPVM 121
131 ARPALNANRWGSLYDALYGTDAIPETDGAEGKPTYNKVRGDKVLAYARKEFLDDSVPLS 190
122 ARPALNANRWGSLYDALYGTDAIPESDGAEGKSYNPKRGEKVJAWRDFLDTBAPLQ 181
191 SGSGFDATGTVDQGLVWALPKSGTGLANPGQ- FAGYTGAASPTSVLLINHLGHEIL 249
182 DCRMKDVGSAFKDGLVRSIDGQAMLTGDKHGFAGYRGDAAAPTHILLKNGIHIEIV 241
250 IDPESQVGTTRDRAGVKVILLESATTTMDPDSVAADKVLGYRNWGLKNGDLAAA 309
242 IDAATTIGKADSAHISDWLESATTTMDCEDSIAVDAEDKVVVYRNWGLKNGDLQEE 301
310 VDKGTAFRLVNRDRNTYAPGGQFTLPGRSLMFRVNVGHMTNDIAVDTDSEVFEPI 369
302 VAKGTSFIRTLNPDLOYAGPDGAAFVHRSMLVENVGHMTNPAILDGRDVEPI 361
370 MDALFTGLIAHGLKASDVNGPLNSRTGSIYIVKPKMGHGAFAVTCFLPSRVEDVL 429
362 MDAATIGLIALYDIPGS---GRRKNSRTGSIYIVKPKMGHGAFAVTCFLPSRVEDALGL 418
430 PONTMKIGIMDEERTTIVNLKACIKAAADRVVINTGFLDRTGDEIHTSMEAGPMVRKGT 489
419 PRNTMKIGIMDEERTTIVNLKESIRAAKDRVVFINTGFLDRTGDEIHTSMEAGPMVRKGT 478
490 MKSOFWILAYEDHNVDAAGLAFSGRAQVKGKMTMELMADVETKIAQPRAGASTAW 549
479 MRQAASISAVENWNVVDIGLECGLSHAQIGKGMWMPDLMAAMLEQKIAHPKAGANTA 538
550 PSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIEEVDN 609
539 PSPTAATLHALHYHVDVAAVQOGLAGKRRATIEQLLTIPLAKELAWAPOBIEEVDN 598
610 QSILGYVVRWVDQVGCCKVDPIDHVALMEDRATLRISSQLLANWLRHGVITSADVRA 669
599 QGILGYVVRWVDQVGCCKVDPIDHVALMEDRATLRISSQLLANWLRHGVITSADVRA 658
670 ERMAPLVDRQVAGDVAYRPMAPNFDSDIARLAAQELILSGAQPNQGYTEPILHRRREFK 729
659 RMAAVVDRQVAGDVAYRPMAPNFDSDIARLAAQELILSGAQPNQGYTEPILHRRREFK 718
730 ARAA 733
719 AKQA 722

RESULT 6
Q92TA4 PRELIMINARY; PRT; 723 AA.
ID AC Q92TA4
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable malate synthase G protein (EC 4.1.3.2).
GN GLCB OR R00062 OR SMC02581.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

```


RESULT 7
Q8YIR3

Db 547 SPTAATLHATHYHKIDVAAVQEKLSRPRAKLDDILSVFVAVRPNTWPTDDIQHEIDNNAQ 606

QY 611 SILGYVVRWDQGVGSKVPDIHDVALMEDRATLRISOLLANLWLRHGVITSADYRASLE 670

Db 607 GILGYVVRWDQGVGSKVPDINNVLGMDRATLRISAQHIANLWLVHGVVSEAQVWETWK 566

QY 671 RMAPLVDQRNAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYVTEPILHRRREFK 729

Db 667 RMAAIVDKQEGDPLRYRPMADFKSIAPQACDLVFKGREQQNGYVTEPVLHRRLELK 725

RESULT 8

Q8FZ50 PRELIMINARY; PRT; 728 AA.

AC Q8FZ50; (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Malate synthase G.

GN GNCB TaxID=29461;

OS Brucella suis.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29461;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1330 / Biovar 1;

RC MEDLINE=22247741; PubMed=122711122;

RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,

RA Read T.D., Dodson R.J., Unayam L., Brinkac L.M., Beanan M.J.,

RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,

RA Nelson W.C., Ayodeji B., Kraul M., Sherry J., Mallek J., Van Aken S.E.,

RA Redmiller S., Fetscherin H., Gill S.R., White O., Salzberg S.L.,

RA Hoover D.B., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;

RT "The Brucella suis genome reveals fundamental similarities between

animal and plant pathogens and symbionts."

RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).

DR EMBL; AE014458; AAN30550.1; -.

DR TIGR; BR1648; -.

KW Complete proteome.

SQ SEQUENCE 728 AA; 79966 MW; F95669D002A14EDE CRC64;

Query Match 38.2%; Score 2313; DB 16; Length 728;

Best Local Similarity 61.6%; Pred. No. 1.3e-125;

Matches 443; Conservative 97; Mismatches 175; Indels 4; Gaps 2;

QY 12 VSVGNLRIARVLYDFVNEALPGTDDIPDSFWAGVGVKVVADLTPOQALLNARDELQAI 71

Db 10 VEIEGLAVPELVFLAKEAPGTGVEPEKFWGFAAIIIRDLAPKRALAKEDELQARI 69

QY 72 DKHRRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPVNA 131

Db 70 DAWKENEDKYSADYQQLKQIGYLLPEGGAFTSVSTNVDPETHIAGPQLVWPVNA 129

QY 132 RFALNAANARWGSYDALYGTVDVTPETDGAKEGPTVNVKRGDKVIAVARKFLDSDVPLS 191

Db 130 RYALNAANARWGSYDALYGTDAISEADGAEGKGYNPKRGEKVIWAKNFDLSAPLST 189

QY 192 GSGDGTGFTVQDQGVVLPDKS-TGLANPGOPAGVTGAESPTSVLLINHLGHIEILI 250

Db 190 GKWADVAGLVNDKLEIRLTDGSAITLKDESQFKYNGDAAFTNVLLAKHNMHVDIVI 249

QY 251 DPESQVGTDRAGVKDVILESATITIMDFSDVAADAADKVLGYRNWLGNKGLDAAV 310

Db 250 NADRPIGKTDPAHITADVIVLESATISTIQCESDIAAADAEDKVAAYRNWLGNKGLDFT 309

QY 311 DKDGTAFVLNRDRNTYAPGGQFTLPGRSLMFRVNVGHMTNDIVDTDGSSEVEGIM 370

Db 310 EKNQKQMTNRUNGORTYAPDGGSTLLTKGHSMLVRNVGHMTNPNPALLDAGNEVEGIM 369

QY 371 DALFTGLIAHLKASDVNGFLINSRTGSIYIVKPKMGHPAEVAFTELSRVEDVGLGP 430

Db 370 DAAFTSLIALHDI--GPNRGRHMSREGSVIVKPKMGHPAEVAFTELSRVEDVGLGP 426

QY 431 QNTMKIGIMDEBERRTTVNLKACIKAAADRVVVFINTGFLDRTGDEIHTSMEAGPMVRKGTM 490

Db 427 PNTLKIGIMDEBERRTTVNLKAEIRAAKDRVVVFINTGFLDRTGDEIHTSMEAGPMIRKGD 486

QY 491 KSQFWILAYEDHNVYDAGLAAGFSGAQVQGVGMWMTLMADMYETKTAQPRAGASTAWVP 550

Db 487 KOAWIGAYEOWNVYDGLGECCLSGHAQIGKGMWMPDMMANLEQKTAHFKAGANTAWVP 546

QY 551 SPTAATLHATHYHGVHVDVAAVQGVQGLAGRRRTIEOLLTIPLAKELAMAPDREEDVNNCQ 610

Db 547 SPTAATLHATHYHKIDVAAVQEKLSRPRAKLDDILSVFVAVRPNTWPTDDIQHEIDNNAQ 606

QY 611 SILGYVVRWDQGVGSKVPDIHDVALMEDRATLRISOLLANLWLRHGVITSADYRASLE 670

Db 607 GILGYVVRWDQGVGSKVPDINNVLGMDRATLRISAQHIANLWLVHGVVSEAQVWETWK 666

QY 671 RMAPLVDQRNAGDVAYRPMAPNFDDSIAPLAAQELILSGAQQNGYVTEPILHRRREFK 729

Db 667 RMAAIVDKQEGDPLRYRPMADFKSIAPQACDLVFKGREQQNGYVTEPVLHRRLELK 725

RESULT 9

Q8FNB3 PRELIMINARY; PRT; 748 AA.

AC Q8FNB3; (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Malate synthase (EC 4.1.3.2).

GN MAS2 OR CE2231.

OS Corynebacterium efficiens.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.

OX NCBI_TaxID=152794;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,

RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,

RA Usuda Y., Sugimoto S.;

RT "The entire genomic sequence of Corynebacterium efficiens YS-314."

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP005221; BAC19041.1; -.

KW Lyase; Complete proteome.

SQ SEQUENCE 748 AA; 83491 MW; F9550473BC4E9A09 CRC64;

Query Match 37.5%; Score 2273; DB 16; Length 748;

Best Local Similarity 61.5%; Pred. No. 2.9e-123;

Matches 444; Conservative 90; Mismatches 182; Indels 6; Gaps 4;

QY 9 TDRVSVGNLRIARVLYDFVNEALPGTDDIPDSFWAGVGVKVVADLTPOQALLNARDELQ 68

Db 26 TERTVVGGMQVAKVLDLFTESVLPVGVDAERFWNGFGDIVRDWTPRNRELLARDELQ 85

QY 69 AQIDKWHRRVIEPIDMDAYRQFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPQLVWPV 128

Db 86 AQLODSEYVNEPKP-DPEKYEAFLEAFLGYLVDPEAPAEIRTONIDSEIATTAGPQLVVP 144

QY 129 LNARFALNAANARWGSYDALYGTVDVTPETDGAKEGPTVNVKRGDKVIAVARKFLDSDV 188

Db 145 LNARFALNAANARWGSYDALYGTNAIPDEDAERGAEPYVPRQKVIQNGRDELDAVLP 204

QY 189 JSSGSGFGDATGFTVQDQGVVLPDKS-TGLANPGOPAGVTGAESPTSVLLINHLGHIEI 248

Db 205 LDGASHADVEXYNTDGLAAHVNVDGIYRLKDRDAYLGTGYTFDPTSIQQNNGHIEL 264

QY 249 LIDPESQVGTDRAGVKDVILESATITIMDFSDVAADAADKVLGYRNWLGNKGLDAA 308

Db 265 QIDPHTPIGKEDKTKGDKIILLESATITIMDFSDVAADAEDKTLGYRNWFLNTGELTE 324

QY 309 AVDKDGTAFVLNRDRNTYAPGGQFTLPGRSLMFRVNVGHMTNDIVDTDGSSEVEG 368

```
Db 325 EVAKGDRFTTRKLNDRVFIGKNGAELTHGKSLIFVRNVGHLMTNPAIL-VDGEEIYEG 383
Qy 369 IMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKQHGPAEVAFTCELFSEVEDVLG 428
Db 384 IMDALIITVCAIPGAPQKKK---NSRKGSIYIVKPKQHGPEEVAFTNELFAEVEDLLD 440
Qy 429 LPQNTMKIGIMDEERTTNVNLKACIKAAADRWVFINTGFLDRTGDEIHTSMEAGPMVRKG 488
Db 441 LPRHTLKGVMDERTSVNLDACIMEVADRLAFINTGFLDRTGDEIHTSMEAGAVRKA 500
Qy 489 TKMSOPWILAYEDHNVDAGLAGSGRAQVKGKGMWMTMELMADWVETKIAOPRAGASTAW 549
Db 501 DMQTAFWQAYEDNNVDAGIQRGUPGKAQIGKGMWMTMELMGEMLKKIGIQLEGANTAW 560
Qy 549 VPSPTAATLHALVHVDVAAVQOGL-AGKGRATEQLLTTIPLAKELAWAPDEIREEVDN 607
Db 561 VPSPTGATLHATHYHRVDVFKVQDELRAAGRDSLSGLKLDVPVAPDNTWTDAREELDN 620
Qy 608 NCQSTILGVVVRWVGCGVCKVPDIHDVALMEDRATLRISQLLANLWLRHGVITSADVRA 667
Db 621 NCQSTILGVVVRWVEGCGVCKVPDIHDIDLMDRATLRISQILANLWLRHGVITSEOVIE 680
Qy 668 SLERNAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAOPNGYTEPIHLHRRRE 727
Db 681 SLERNVVDVDRQAGDPNLYMAPNFTESVAFQARDLILKGTSPAGYITEPIHLHRRRE 740
Qy 728 FK 729
Db 741 FK 742

RESULT 10
Q9KB03 PRELIMINARY; PRT; 727 AA.
ID Q9KB03
AC Q9KB03
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Malate synthase.
GN BH2133.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL; AP001514; BAB05852.1; -
DR HSSP; P37330; 1D8C.
DR InterPro; IPR001465; Malate synthase.
DR InterPro; IPR006253; Malate synthase.
DR Pfam; PF01274; Malate synthase; 1.
DR TIGRfams; TIGR01345; malate_syn_G; 1.
SQ Complete proteome.
QY SEQUENCE 727 AA; 81208 MW; CE18B45930E5EAE5 CRC64;

Query Match
Best Local Similarity 37.3%; Score 2258; DB 16; Length 727;
Matches 441; Conservative 97; Mismatches 166; Indels 12; Gaps 4;

Qy 1 MOHHHHHTDRVSVGNLRIARVLYDFVNEALPGTDIDPDPSFWAGVDKVKVADLTFQNAL 60
Db 1 METYEH-----VGNLQVATKLITFIEQRALPGTGKIKDDFWSGVQQLITELMPENKML 53
Qy 61 LNARDELQADIKVHRRAVIEPIDMDAYRQFLTEIGYLLPEPDFTTITTSVDRIITTA 120
Db 54 LAKREEIQATIDAMHQKNK-GPIDFSAYHSFLEIGYLEP:PEHVITITENVVDDEIAQA 112
```

```
Qy 121 GPOLVVPVNLNARPALNANRWGSLYDALYGTDVIPETDGAEGKPTYNKVRGDKVIAYAR 180
Db 113 GPOLVVPVNNARVAINANRWGSLYDALYGSNVISEEDGCEKGTGYNPKRGTKVIOFAK 172
Qy 181 KFDDSDVPSGSGFGDATGFTVQDQGLVVALPD-KSTGLANPQCFAGYTCGAESPTSVLL 239
Db 173 DFLDHTFPLTSGSHHEAINVAKMDQLVLTESGKMTLKDFTQFVYGSGSQDPSVILL 232
Qy 240 INHGLHIELIDPESOGVITDRAGVKDVLKESAITTIMDFEDSVAAADKVLGVNWL 299
Db 233 LHGLHVEIQIDARHPIGKSDRAKVDIVLESALITTIMDCEDSVAAADKEDKVAVYQNL 292
Qy 300 GLNKGDLAAAVDKGTAFLVLRNDRNYTAPGGQFTLPGRSLMFRVNVGHMLTNDIVD 359
Db 293 GLMKGTLKATFKEGTKKKNEDRSYTAPEGETFSLGRSLMFRVNVGHMLTTPVIRT 352
Qy 360 TDGSEVFEGIMDALFTGLIAIHGLKASDVNGPLNSRTGSIYIVKPKQHGPAEVAFTCEL 419
Db 353 QSSEVEPEGLDGIIVTSLIAKHDLQ---NGTFRNSKGSVIVKPKHGSBEVAFANRL 409
Qy 420 FSRVEDVLGLPQNTMKIGIMDEERTTNVNLKACIKAAADRWVFINTGFLDRTGDEIHTSM 479
Db 410 FNRIEDILGLERTLKIWDERRISLNLKACIEKVKERVVFINTGFLDRTGDEIHTSM 469
Qy 480 EAGPMVRKGTMSQPMILAYEDHNVDAGLAGSGRAQVKGKGMWMTMELMADWVETKIAQ 539
Db 470 EAGPMIRKGMKSSWLSAYERSNVAAGLTCGFGRAQIGKGMWAMPDLNMEWMEQKGTQ 529
Qy 540 PRAGASTAWPSPATAATHALVHVDVAAVQOGLAGKGRATEQLLTTIPLAKELAWAPD 599
Db 530 LEAGANTAWPSPATAATHALVHRRHVPAIQKTLADDQTCYRKKEILEIPLAGSTDTWNE 589
Qy 600 EIREEDVNNQSTILGVVVRWVGCGVCKVPDIHDVALMEDRATLRISQLLANLWLRHGV 659
Db 590 DIQAEELNNAQGLIGYVVRWVEGCGVCKVPDIHNTALMEDRATLRISQHANLWLRHGI 649
Qy 660 ITSADVRSASLERMAPLVDRQAGDVAYRMAPNFDSDIAFLAAQELILSGAOPNGYTEP 719
Db 650 VSXEQVIRTMERMAKVVDQAGDPAYRPMADNLEQSVAFQAALEVLKGTQEPSGYTEP 709
Qy 720 ILHRRRREFKARAAEK 735
Db 710 ILHRRLEFKQIAKE 725

RESULT 11
Q8FDN6 PRELIMINARY; PRT; 723 AA.
ID Q8FDN6
AC Q8FDN6
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Malate synthase G (EC 4.1.3.2).
GN GUCB OR C3705.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016766; AA82151.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 723 AA; 80440 MW; AAF740B5FE038F6F CRC64;
```

Query Match 35.8%; Score 2168.5; DB 16; Length 723;
Best Local Similarity 58.4%; Pred. No. 3.1e-117; Indels 15; Gaps 5;
Matches 422; Conservative 113; Mismatches 173;

QY 12 VSVNGLRIARVLVDVFNNEALPGTDIDPDSFWAGVDKVVADLTTPQNCALLNARDELQAOI 71
DB 5 ITQGLRIDANPKRFVDEEVLPGVELDAAAFWHNVDEIVHDLAPENRQLLAERDRIQAAL 64

QY 72 DKHRRVRIEPIIDMAYRFLTEIGYLLPEPDDFTITTSQVDAEITTTAGPOLVVPVNA 131
DB 65 DEWHRNPGSVKDAAYKFLRELGLVLPQPDHVTVEITGIDSEITSQAGPOLVVPVNA 124

QY 132 RFLNANANRWGLYDALYGTVDVPIPTDGAEGKPTYNKVRGDKVAYARKPDDSDVPLSS 191
DB 125 RYALNANANRWGLYDALYGSIIPOEGAVSG--YDPQGEQVIAWVRFIDESLPLEN 182

QY 192 GSGDATFTVQDGLVVALPD-KSTGLANPGQFAGYTGAAESPTSVLLINHLHIEILI 250
DB 183 GSYQDVVAFKVDKQLRIQLKNGKETTLRTPAQFVGYRGDTAAPTICILLKXNGLHIEIQI 242

QY 251 DPSSQVGTTRAGVKDVILESATTTIMDPEDSVAADKVLGYRNWGLNKGDLAAAV 310
DB 243 DANGRIKGDASAHINDVIVEAALSTIDCEDSVAAVDAEDKILLRNLLGLMQGTLOEKX 302

QY 311 DKDGTAFRLVNRDRNTAPGGQFTLPGRLMFVRNVGHMTNDAIVDTDGEVPEGIM 370
DB 303 EKNGRQIVRKLNDRDQVTAADGSEISLHGRSLFIRNVGHLMTIPVIMDSQEGNEIPEGIL 362

QY 371 DALFTGLIAHGLKASDVNGPLNSRSTGYIVKPKMGPAEVAFTCELFSEVEDVLGLP 430
DB 363 DGVTMGALYDLKVK-----NSRTGSYIVKPKMGPEQVAFANKLFSRVETMLGNA 416

QY 431 QNTWKIGIMDEERTTNLKAICAKAAADRVVINTFGFLDRTGDEIHTSMEAGPMVRKGTM 490
DB 417 PNLKXGIMDEERTSINLRSCIAQARNVAFINTGFLDRTGDEMHSVVEAGPMLRKQM 476

QY 491 KSPWILAYEDHNVDAAGLAFSGRAQVKGKMTWELMADWVETKIAOPRAGASTAWVP 550
DB 477 KSTPWIKAYERNVLSGLFCGLRGKAQIGKGMWMPDLMDMYSQKGDQLRAGANTAWVP 536

QY 551 SPTAATLHALHYQVDVAAVVOOGLA-----GKRATIEOLLTIPLAKELAWAPDEIREV 605
DB 537 SPTAATLHALHYQTNVQSVQANIAQTEFNAEPEPLLDLLITPVAENANNSVEHILQEL 596

QY 606 DNNQSTILGVYVRVQGVGCKVPDIHVDMEDRATLRISQLLANWLHGVITSADV 665
DB 597 DNNVQIGLVYVRVWVEQIGCKSVKVPDIHVDMEDRATLRISQHLANLHGHILTKEQV 656

QY 666 RASLERMAPVDONAGDVAYRPMANFDDSIAPLAAQELILSGAQQNGYTEPILHR-R 724
DB 657 QASLENKAVVDQONAGDPAYRPMAGNFANSCAFKASDLIFLGVKQPNGYTEPILHAWR 716

725 RRE 727
717 LRE 719

RESULT 12

Q9CB77 PRELIMINARY; PRT; 421 AA.
ID Q9CB77
AC Q9CB77
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Aspartokinase.
GN ASK OR ML2323.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.

RESULT 13

Q8KQ29 PRELIMINARY; PRT; 421 AA.
ID Q8KQ29
AC Q8KQ29
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Aspartate kinase alpha subunit (EC 2.7.2.4) (Aspartokinase).
GN ASKA.
OS Streptomyces clavuligerus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1901;
RN [1]
RP SEQUENCE FROM N.A.

Tunca S';
RA "Isolation and characterization of the aspartokinase and aspartate
RT semialdehyde dehydrogenase genes from cephamycin C-producer
RT Streptomyces clavuligerus";
RL Thesis (2002), University of Middle East Technical University, Ankara,
Turkey.
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL: AV112728; AM54736.1; -.
DR InterPro: IPR001048; Asp_kinase.
DR InterPro: IPR002912; ACT.
DR InterPro: IPR001341; Aspartate_kinase.
DR InterPro: IPR005260; Asp_kin_monofn.
DR InterPro: IPR001057; Glu_Skinase.
DR Pfam: PF00696; aak_kinase; 1.
DR Pfam: PF01842; ACT; 2.
DR PRINTS: PRO0474; GLUSKINASE.
DR TIGRFAMs: TIGR00657; asp_kinases; 1.
DR TIGRFAMs: TIGR00656; asp_kin_monofn; 1.
DR PROSITE: PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44399 MW; 08262D81045735C1 CRC64;

Query Match 21.0%; Score 1272.5; DB 2; Length 421;
Best Local Similarity 66.1%; Pred. No. 1.1e-65;
Matches 237; Conservative 51; Mismatches 60; Indels 21; Gaps 4;

QY 749 VQKYGSSVADAEIRIRVAERIVATKQGNVNVVVSAMGDTTDLDDLAQVCVPPPPR 808
DB 5 VQKYGSSLESADRIKVAERIVATKAGNEVVVVCVSGMDTTDELLDQAQVNPVPPER 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAIQAQAWSFSGAGVITTVAVHGNARIIDVTPSRVSEA 124

QY 869 LEEGRVVLVAGFGVSGQDTKVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928
DB 125 LEEGVVALVAGFGVAGQADTKITTLGRGSDTTAVALLAALNADVCETYSVDVGYSADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYAREHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVPDARKLDSVPEEMLELAASGSKILHLSRVEYARYGVPIRVSSYSKPGTTVTGS 244

QY 989 IKDVPMEPDLITGVADHRSSEAKVTIVGLPDIPGYAAKVFAVARRRRQHRHGAEE 1043
DB 245 IEEIPVEQALITGVADHRSSEAKITVGVDPDHAGAAARIFRIVA-----DAEIDDM 295

QY 1044 RLQG-----GRGQDRHLLHLLPOTSGPPPKWTRSETRSASTOLLYDDHIGKVSLLGAG 1098
DB 296 VLQVNSVTSVSGRTD--ITFTLSKANGKAVASLEKIEELGFSVLYDDHVGKVSLLVGAG 353

QY 1099 MRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 MRSHPGVTATFCEALSKVGVNIEINTSE 382

RESULT 14
Q9RQ25 PRELIMINARY; PRT; 421 AA.
AC Q9RQ25;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 23, Last annotation update)
DE Aspartokinase subunit A (EC 2.7.2.4) (Aspartate kinase).
GN ASKA.
OS Amycolatopsis mediterranei (Nocardia mediterranei).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardiaaceae; Amycolatopsis.
OX NCBI_TaxID=33910;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U32;

MEDLINE=99453302; PubMed=10521665;
RA Zhang W., Jiang W., Zhao G., Yang Y., Chiao J.;
RT "Sequence analysis and expression of the aspartokinase and aspartate
RT semialdehyde dehydrogenase operon from rifamycin SV-producing
RT amycolatopsis mediterranei";
RL Gene 237:413-419(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC EMBL: AF134837; AAD49567.1; -.
DR InterPro: IPR001048; Asp_kinase.
DR InterPro: IPR002912; ACT.
DR InterPro: IPR001341; Aspartate_kinase.
DR InterPro: IPR005260; Asp_kin_monofn.
DR InterPro: IPR001057; Glu_Skinase.
DR Pfam: PF00696; aak_kinase; 1.
DR Pfam: PF01842; ACT; 2.
DR PRINTS: PRO0474; GLUSKINASE.
DR TIGRFAMs: TIGR00657; asp_kinases; 1.
DR TIGRFAMs: TIGR00656; asp_kin_monofn; 1.
DR PROSITE: PS00324; ASPARTOKINASE; 1.
KW Kinase; Transferase.
SQ SEQUENCE 421 AA; 44393 MW; 633D9C2D023145E9 CRC64;

Query Match 20.9%; Score 1267.5; DB 2; Length 421;
Best Local Similarity 65.3%; Pred. No. 2.2e-65;
Matches 254; Conservative 53; Mismatches 61; Indels 21; Gaps 4;

QY 749 VQKYGSSVADAEIRIRVAERIVATKQGNVNVVVSAMGDTTDLDDLAQVCVPPPPR 808
DB 5 VQKYGSSLESADRIKVAERIVATKAGNNDVVVVCVSGMDTTDELLDQAQVNPAPPER 64

QY 809 ELDMLLTAGERISNALVAMAIESLGAHARSFTGSGAGVITTTGHNAKIIDVTPGRLQTA 868
DB 65 EMDMLLTAGERISNALVAMAIQAQAWSFSGAGVITTVAVHGNARIIDVTPSRVTEA 124

QY 869 LEEGRVVLVAGFGVSGQDTKVTTLGRGSDTTAVAMAAALGADVCEIYTDVDFISADP 928
DB 125 LDQYIALVAGFGVAGQADTKITTLGRGSDTTAVALLAALNADVCETYSVDVGVTADP 184

QY 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYAREHNPVHVRSSYSDRPGTVVVG 988
DB 185 RIVPDARKLDSVPEEMLELAASGSKILHLSRVEYARYGVPIRVSSYSKPGTTVTGS 244

QY 989 IKDVPMEPDLITGVADHRSSEAKVTIVGLPDIPGYAAKVFAVARRRRQHRHGAEE 1043
DB 245 IEEIPVEQALITGVADHRSSEAKITVGVDPDHAGAAARIFRIVA-----DAEIDDM 295

QY 1044 RLQG-----RGQDRHLLHLLPOTSGPPPKWTRSETRSASTOLLYDDHIGKVSLLGAG 1098
DB 296 VLQVNSVTSVSGRTD--ITFTLSKANGKAVKELEKVAEIGFESVLYDDHVGKVSLLVGAG 353

QY 1099 MRSHPGVTATFCEALAAVGVNIELISTSE 1127
DB 354 MRSHPGVTATFCEALAEAGVNIIEINTSE 382

RESULT 15
Q93C57 PRELIMINARY; PRT; 421 AA.
AC Q93C57;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Feedback-resistant aspartokinase LysC alpha subunit (EC 2.7.2.4)
(Aspartate kinase).
OS Corynebacterium crenatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=168810;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD945;

RA Liu Y., Ding J., Wang J.;
RT "Cloning and sequence analysis of aspartokinase genes from
RT *Cornebacterium crenatum*,"
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
CC -/- CATALYTIC ACTIVITY: ATP + L-ASPARTATE = ADP + 4-PHOSPHO-L-
CC ASPARTATE.
CC -/- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.

DR EMBL; AF414084; AAL07807.1; -.

DR InterPro; IPR001048; Aa_kinase.

DR InterPro; IPR002912; ACT.

DR InterPro; IPR001341; Aspartate_kinase.

DR InterPro; IPR005260; Asp_kin_monofn.

DR Pfam; PF00696; aakinese; 1.

DR Pfam; PF01842; ACT; 2.

DR TIGRFAMS; TIGR00657; asp_kinases; 1.

DR TIGRFAMS; TIGR00656; asp_kin_monofn; 1.

DR PROSITE; PS00324; ASPARTOKINASE; 1.

KN Kinase; Transferase.

SQ SEQUENCE 421 AA; 44738 MW; 473A19409C0215E6 CRC64;

Query Match 20.3%; Score 1232.5; DB 2; Length 421;

Best Local Similarity 64.3%; Pred. No. 2.3e-63;

Matches 252; Conservative 55; Mismatches 58; Indels 27; Gaps 5;

Qy 749 VQYGGSSVADAEIRIRVAERIVATKQGNVWVVSAMGDTTDDLDAQQVCAPPER 808

Db 5 VQYGGSSLESARIRVAERIVATKAGNDVVVCSAMGDTTDELEAAAVNPVPPAR 64

Qy 809 ELDMLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLQTA 868

Db 65 EMDMLTAGERISNALVAMAEISLGAHARSFTGSQAGVITTTGTHGNAKIIVTPGRLQTA 124

Qy 869 LEEGRVVLVAGFQGVSOQTDVTTLGRGSDTTAVAMAALGADVCEIYTDVDCIFSDP 928

Db 125 LDEGKICIVAGFQGVNKTEDVTTLGRGSDTTAVALAALNADVCEIYSDVDGVYADP 184

Qy 929 RIVRNARKLDTVTFEEMLEMAACGAKVLMRCVEYARRHNI PVHVRSSYSRPGTVVWGS 988

Db 185 RIVPNAQKLEKLFEEMLELAARVGSKILVLRSEYARAFNPLRVSSYSNDPGTILAGS 244

Qy 989 IKDVPWEDPILTVADHRSKVTIVGLPDIPGYAAKVFRVA- 1038

Db 245 MEDIPVEEAVLTGVATDKSEAKVTVLGSDKPGAAKVFRALADAEINIDMWLVNVSSYE 304

Qy 1039 HGAERLQGRGROD-RHHLHL--POTSGPPPKNWTSETRSASTQLLYDDHIGKVSLL 1095

Db 305 DGTDTITFTCPRSDGRAMEILKKPQVQG-----NW-----TNVLYDDGVGVSLV 350

Qy 1096 GAGMRSHPGVTATFCEALAAVGNIELISTSE 1127

Db 351 GAGWKSHPGVTABFMEALRDVNVNIELISTSE 382

Search completed: November 21, 2003, 16:08:15
Job time : 60.6698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 59.515 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHVHHVHVIDIGTPTSW.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5072	100.0	983	22 AAU01901	M. tuberculosis Tb
2	2700	53.2	802	19 AA081746	M. tuberculosis fu
3	2700	53.2	802	19 AA064379	Mycobacterium anti
4	2700	53.2	802	20 AA032063	Mycobacterium tube
5	2700	53.2	802	20 AA039224	M. tuberculosis fu
6	2700	53.2	802	20 AA039224	M. tuberculosis fu
7	2700	53.2	802	20 AA039081	M. tuberculosis fu
8	2700	53.2	802	20 AA039033	M. tuberculosis fu
9	2700	53.2	802	23 AAU74592	Antigenic fusion p

10	2375	46.8	460	19 AA081737	M. tuberculosis im
11	2375	46.8	460	19 AA064367	Mycobacterium tube
12	2375	46.8	460	20 AA039164	M. tuberculosis re
13	2375	46.8	460	20 AA039021	M. tuberculosis re
14	1839	36.3	652	20 AA039225	M. tuberculosis fu
15	1839	36.3	652	20 AA039082	M. tuberculosis fu
16	1831	36.1	374	19 AA081730	M. tuberculosis fu
17	1831	36.1	374	19 AA081731	M. tuberculosis 38
18	1831	36.1	374	19 AA064363	Mycobacterium anti
19	1831	36.1	374	19 AA064364	Mycobacterium 39 k
20	1831	36.1	374	20 AA032061	Mycobacterium tube
21	1831	36.1	374	20 AA039161	M. tuberculosis an
22	1831	36.1	374	20 AA039160	M. tuberculosis Tb
23	1831	36.1	374	20 AA039017	M. tuberculosis fu
24	1831	36.1	374	20 AA039018	M. tuberculosis 38
25	1831	36.1	374	22 AA011840	Mycobacterium tube
26	1831	36.1	374	23 AA029719	Mycobacterium sp.
27	1831	36.1	374	23 AA074590	Antigenic fusion p
28	1831	36.1	374	23 AA017593	Mycobacterium spec
29	1831	36.1	374	23 AA050733	Mycobacterium tube
30	1808	35.6	374	13 AA030090	38 kDa lipoprotein
31	791	15.6	166	18 AA032445	Mycobacterium tube
32	791	15.6	166	18 AA032377	Mycobacterium tube
33	791	15.6	166	19 AA081681	M. tuberculosis im
34	791	15.6	166	19 AA064319	Mycobacterium tube
35	791	15.6	166	20 AA039119	M. tuberculosis an
36	791	15.6	166	20 AA038982	M. tuberculosis re
37	485	9.6	100	19 AA081706	M. tuberculosis im
38	485	9.6	100	19 AA064339	Mycobacterium tube
39	485	9.6	100	20 AA039136	M. tuberculosis an
40	485	9.6	100	20 AA038993	M. tuberculosis re
41	485	9.6	100	20 AA035218	M. tuberculosis Rv3
42	485	9.6	100	22 AA019845	Mycobacterium tube
43	485	9.6	100	23 AB005988	M. tuberculosis an
44	485	9.6	100	23 AB030969	Mycobacterium tube
45	481	9.5	100	20 AA03705	M. tuberculosis LH

ALIGNMENTS

RESULT 1

AAU01901
ID AAU01901 standard; Protein; 983 AA.

XX AAU01901;

AC AAU01901;

XX 29-AUG-2001 (first entry)

DT M. tuberculosis TbF15 fusion protein.

XX M. tuberculosis TbF15 fusion protein.

DE TbF15; antigen; vaccine; tuberculosis; AIDS; His tag; Ra3; 38kD;

XX 38-1; FL TbH4; acquired immunodeficiency disease.

KW Synthetic.

XX Mycobacterium tuberculosis.

OS Mycobacterium tuberculosis.

XX Key Binding-site

FT Location/Qualifiers

FT 3..8

FT /label= Histidine tag

FT /note= "Nickel chelating region used for purifying

FT the fusion protein"

FT 9..74

FT /label= Ra3_region

FT 75..425

FT /label= 38kD_region

FT 426..524

FT /label= 38-1_region

FT 525..983

FT /label= FL_TbH4_region

XX WO200124820-A1.

XX

CC protein is used in a method for inducing protective immunity against
CC tuberculosis (TB). This sequence can be formulated into vaccines
CC and/or pharmaceutical compositions for immunising against
CC M. tuberculosis infection or may be used for the diagnosis of TB.
XX
SQ Sequence 802 AA;
Query Match 53.2%; Score 2700; DB 19; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTLLYPLFLNWG 120
Db 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTLLYPLFLNWG 119
QY 121 PAFHERYPNVTITAGTSGAGIAQAAAGTNYIGASDAYLSEGDMAAHKLGNIALAISA 180
Db 120 PAFHERYPNVTITAGTSGAGIAQAAAGTNYIGASDAYLSEGDMAAHKLGNIALAISA 179
QY 181 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
QY 241 GDTFLFTQYLSKODPEGKGPFGFTTVDFFAVPGALGENGGWVTGCAETPCGVAYIG 300
Db 240 GDTFLFTQYLSKODPEGKGPFGFTTVDFFAVPGALGENGGWVTGCAETPCGVAYIG 299
QY 301 ISFLDQASQRLGEAQLGNSSGNFLFLPDQAQIAAAAGFASKTPANQAISMIDGPA PDGY 360
Db 300 ISFLDQASQRLGEAQLGNSSGNFLFLPDQAQIAAAAGFASKTPANQAISMIDGPA PDGY 359
QY 361 PIINVEYAIVNRRQKDAATQALQFLHWAITDGNKASFLDQVHPQPLPRAVVKLS DALI 420
Db 360 PIINVEYAIVNRRQKDAATQALQFLHWAITDGNKASFLDQVHPQPLPRAVVKLS DALI 419
QY 421 ATISAEAKTKDAATLAQAGNFERISGDLTKQIDQVESTAGSLQOQWGAAGTAAQA AVV 480
Db 420 ATISAEAKTKDAATLAQAGNFERISGDLTKQIDQVESTAGSLQOQWGAAGTAAQA AVV 479
QY 481 RFQEAANKQKQLDISINIRQAGVQYGRABEEQQOALSSOMGFTQSTVTVDDQOEILNR 540
Db 480 RFQEAANKQKQLDISINIRQAGVQYGRABEEQQOALSSOMGFTQSTVTVDDQOEILNR 537
QY 541 ANEVEAPNADPTDVPITPCELTAAKNAQQVLSDNNREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATPVAPPPAAANTPNQPGDPNAPPPADPNAPPPVIAPNADQPV R----- 589
QY 601 AKAYGEVDEEATALDNDGEGTVQAESAGVGGDS-----SABLTDTPRVATAGEPNF- 653
Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDP PFP 630
QY 654 -----MDLKEARKLETGDOGASLAHEADGWNFTNLTLQGDVXFRFGPD 697
Db 631 GQPPVANDTRIVLGRLOQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEGDAATACEASLDQORQWILHMAKLSAAMAKAQYVAQLFVWARREHPTIEDIVGLER 757
Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPNQKQWTVGIGSPA 724
QY 758 LYAENPSARDQILPVAEYQORSEKVLTEYNN-----KAALEFVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPQFVFW-----LGTANNPVDKGAAKALAESIRLPAVPPA-----P 770
QY 810 PPPPOEQGLIPGFLMPPSDGSGVTGTMGAAPVPPPTGSGGGGLPA 856
Db 771 APAPAEPA-----PAPAPAEVAP-----TPTTFTFORTLPA 802

RESULT 3

AAW64379
ID AAW64379 standard; Protein; 802 AA.
XX
AC AAW64379;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium antigen TbF2 protein fusion.
XX
KW Tuberculosis; infection; diagnosis; 38 kDa antigen; TBra3; DPEP;
KW TB38-1; TbF-2.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
PN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV55801.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 7; Page 223-226; 250pp; English.
XX
CC This polypeptide comprises a fusion protein, designated TbF-2,
CC composed of Mycobacterium tuberculosis antigens Tbra3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. TbF-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression
CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.
SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 19; Length 802;
Best Local Similarity 66.6%; Pred. No. 2.9e-158;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHHEHHVHDIIGTSPTSWEQAAAEAVQRAEDSVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTLLYPLFLNWG 120
Db 61 KLEVSFKMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGSTLLYPLFLNWG 119
QY 121 PAFHERYPNVTITAGTSGAGIAQAAAGTNYIGASDAYLSEGDMAAHKLGNIALAISA 180
Db 120 PAFHERYPNVTITAGTSGAGIAQAAAGTNYIGASDAYLSEGDMAAHKLGNIALAISA 179
QY 181 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

QY 241 GDTFLFTQYLSKODPBGWKGSPGFTTVDFFPAVPGALGNGMGVMTGCAETPGCVAYIG 300
 DB |||||
 DB 240 GDTFLFTQYLSKODPEGWKGSPGFTTVDFFPAVPGALGNGMGVMTGCAETPGCVAYIG 299
 QY 301 ISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMIDGPA PDGY 360
 DB |||||
 DB 300 ISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMIDGPA PDGY 359
 QY 361 PIINYEVAI VNNRQKDAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 420
 DB |||||
 DB 360 PIINYEVAI VNNRQKDAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 419
 QY 421 ATISSAEMKTDAAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 480
 DB |||||
 DB 420 ATISSAEMKTDAAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSOMGFTQSQTIVTVDQOEILNR 540
 DB |||||
 DB 480 RFOEAANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSOMGFTQSQTIVTVDQOEILNR 537
 QY 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 DB |||||
 DB 538 PPAPATVPAPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAQPVPR----- 589
 QY 601 AKAYGEVDEEAATLNDGEGTVOAESAGAVGDS-----SALTDTPRVTAGEPNP- 653
 DB |||||
 DB 590 -----IDNPVGFGFSFALPAGWVESDAAHFDYGSALLS-----KITGDPDPFP 630
 QY 654 -----MDLKEAARKLETGDOGASLAHAFADGWNFTNLTLQGVKFRGFD 697
 DB |||||
 DB 631 GQPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDNGEF--YM 675
 QY 698 NWEQDAATACEASLDQOROMILHNAKLSAAMAKQAOYVAQLHWARREHPTVEDIVGUER 757
 DB |||||
 DB 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDPKPNQIWTGVIGSPA 724
 QY 758 LYAENSARDQIILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 DB |||||
 DB 725 ANAPDAGPQORWFWVW-----LGTANNPVYKGAALAEISRPLVAPPPA-----P 770
 QY 810 PPPPQEQGLIFGLPMPSPSDGSGVTPGTGMPAPAPMPVPTGSPGGPLPA 856
 DB |||||
 DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTTPORTLPA 802

RESULT 4
 AAY32063
 ID AAY32063 standard; Protein; 802 AA.
 AC AAY32063;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein TbF-2.
 XX
 KW Tuberculosis; antigen; fusion protein; TbF-2; Tbra3; 38kD; Tb38-1;
 KW DREP; diagnosis; therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US07717.
 XX
 PR 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Alderson M, Campos-Neto A;

XX WPI; 1999-601610/51.
 DR N-PSDB; AAZ20198.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis -
 XX
 PS Claim 1; Fig 5G-J; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis
 CC tetra-antigen fusion protein, termed TbF-2, composed of the antigens
 CC Tbra3, 39kD, Tb38-1 and DREP. The fusion protein is expressed in
 CC host cells using a vector carrying a polynucleotide (see AAZ20198)
 CC comprising the 4 coding sequences. The invention provides fusion
 CC proteins (see AAY32059-71) containing at least 2 M. tuberculosis
 CC antigens. The new fusion proteins and polynucleotides encoding
 CC them are useful as vaccines for preventing tuberculosis (claimed),
 CC for diagnosis (via in vitro assays or intradermal skin tests for
 CC detection of anti-M. tuberculosis antibodies), monitoring of
 CC disease progression, and treatment of tuberculosis. They are more
 CC effective immunogens than mixtures of the individual protein
 CC components.
 XX
 SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHHVHDIIGTSPTSWEQAAAEEAVQARQSVDDIRVARVIEQDMVDSAGKITRYI 60
 DB |||||
 QY 1 MGHHHHHHVHDIIGTSPTSWEQAAAEEAVQARQSVDDIRVARVIEQDMVDSAGKITRYI 60
 DB |||||
 QY 61 KLEVSFKORPAQPRCGSKPGSPETGAGAGTATTTPASSPVTIETGSLTLLPLFNWLG 120
 DB |||||
 QY 61 KLEVSFKORPAQPR-GSKPPSGSPETGAGAGTATTTPASSPVTIETGSLTLLPLFNWLG 119
 QY 121 PAFHERPVNTITTAQGTGSGAGIAQAAAGTNTIGASDAYLSEGDMAHKLMIALAIISA 180
 DB |||||
 QY 120 PAFHERPVNTITTAQGTGSGAGIAQAAAGTNTIGASDAYLSEGDMAHKLMIALAIISA 179
 DB |||||
 QY 181 QOVNPNLPGVSEHLKNGKVLAAAMYQCTIKTWDDPQIAALNPGVNLPGTAVVFLHRS DGS 240
 DB |||||
 QY 180 QOVNPNLPGVSEHLKNGKVLAAAMYQCTIKTWDDPQIAALNPGVNLPGTAVVFLHRS DGS 239
 QY 241 GDTFLFTQYLSKODPEGWKGSPGFTTVDFFPAVPGALGNGMGVMTGCAETPGCVAYIG 300
 DB |||||
 QY 240 GDTFLFTQYLSKODPEGWKGSPGFTTVDFFPAVPGALGNGMGVMTGCAETPGCVAYIG 299
 QY 301 ISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMIDGPA PDGY 360
 DB |||||
 QY 300 ISFLDQASQRLGEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMIDGPA PDGY 359
 QY 361 PIINYEVAI VNNRQKDAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 420
 DB |||||
 QY 360 PIINYEVAI VNNRQKDAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 419
 QY 421 ATISSAEMKTDAAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 480
 DB |||||
 QY 420 ATISSAEMKTDAAATLQAEAGNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSOMGFTQSQTIVTVDQOEILNR 540
 DB |||||
 QY 480 RFOEAANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSOMGFTQSQTIVTVDQOEILNR 537
 QY 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 DB |||||
 QY 538 PPAPATVPAPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAQPVPR----- 589
 QY 601 AKAYGEVDEEAATLNDGEGTVOAESAGAVGDS-----SALTDTPRVTAGEPNP- 653
 DB |||||
 QY 590 -----IDNPVGFGFSFALPAGWVESDAAHFDYGSALLS-----KITGDPDPFP 630

QY 654 -----MDLKEARKLETGDOGASLAHFADGWNTFNLTQGDVRRGPD 697
 Db 631 GPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEP--YM 675
 QY 698 NWEGDAATACASLDQOOROWILHMAKLSAAMAKQAQYVAQLHVWARRHEHTYEDIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPNQGIWTVIGSPA 724
 QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAESIRPLVAPPPA---P 770
 QY 810 PPPQEOGLIFGLFMPSPDSGVTPTGMPAAPMVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 5
 ID AAY39224
 XX AAY39224 standard; Protein; 802 AA.
 AC AAY39224;
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbF-6 amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527409/44.
 DR N-PSDB; AA219457.
 XX
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Claim 37; Page 271-273; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.
 XX
 SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

DE M. tuberculosis fusion protein TbF-2 amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

QY 1 MGHHHHHVIDIIGTSPTSEQAAAEAVQARSDVDIRVARVIEQDMAVDSAGKITYRI 60
 Db 1 MGHHHHHVIDIIGTSPTSEQAAAEAVQARSDVDIRVARVIEQDMAVDSAGKITYRI 60
 QY 61 KLEVSFKMRPAQPCRCSPSPETGAGAGTVAATPASPVTLAETGTLTYPLENMG 120
 Db 61 KLEVSFKMRPAQPR-GSKPPSGSPETGAGAGTVAATPASPVTLAETGTLTYPLENMG 119
 QY 121 PAFHERYPNVTTAQTGSGAGIAOAAAGTVAIGASDAYLSEGDMAAHKGLNIALAISA 180
 Db 120 PAFHERYPNVTTAQTGSGAGIAOAAAGTVAIGASDAYLSEGDMAAHKGLNIALAISA 179
 QY 181 QQVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSG 240
 Db 180 QQVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSG 239
 QY 241 GDTFLTQYLSKQDPGKSKSPGFGTTVPFPAVPGALGENGNGMTGCAETPGCVAYIG 300
 Db 240 GDTFLTQYLSKQDPGKSKSPGFGTTVPFPAVPGALGENGNGMTGCAETPGCVAYIG 299
 QY 301 ISFLDOASQRLGEAQLGNSSGNFLPDAGSIQAAAAGFASKTPANCAISMIDGPADGY 360
 Db 300 ISFLDOASQRLGEAQLGNSSGNFLPDAGSIQAAAAGFASKTPANCAISMIDGPADGY 359
 QY 361 PIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVYKLSALI 420
 Db 360 PIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVYKLSALI 419
 QY 421 ATTSSAEMKTDAAATLAQEAQNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAVV 480
 Db 420 ATTSSAEMKTDAAATLAQEAQNFERISGDLTKTQIDQVESTAGSLQGWGGAAGTAAQAAVV 479
 QY 481 RFOEAAKQKQELDEISTNIRQAGVQYSRADEQQAALSSQMGFTQSTQTVTDQOELNR 540
 Db 480 RFOEAAKQKQELDEISTNIRQAGVQYSRADEQQAALSSQMGFV--PTTAAAPPSTAAA 537
 QY 541 ANEVEAPMADPTDVTITPCELTAAKNAQQLVLSADNREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATPVAPPPPPAAANTFNAPQDPNAPPPPPADPNAPPPPPVAPNAPQVPR-----589
 QY 601 AKAYGEVDEEAATALDNDGEGTVOAESAGAVGDS-----SALETDTPRVATAGBNF- 653
 Db 590 -----INDPVGFSFALPAGVSDAAHFDYGSALLS-----KITGDPFPP 630
 QY 654 -----MDLKEARKLETGDOGASLAHFADGWNTFNLTQGDVRRGPD 697
 Db 631 GPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEP--YM 675
 QY 698 NWEGDAATACASLDQOOROWILHMAKLSAAMAKQAQYVAQLHVWARRHEHTYEDIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSKPNQGIWTVIGSPA 724
 QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAESIRPLVAPPPA---P 770
 QY 810 PPPQEOGLIFGLFMPSPDSGVTPTGMPAAPMVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 6
 AAY39176
 ID AAY39176 standard; Protein; 802 AA.
 XX
 AC AAY39176;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbF-2 amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

OS Synthetic.
 OS Mycobacterium tuberculosis.

XX WO9942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX PR 05-MAY-1998; 98US-0072967.

XX PR 18-FEB-1998; 98US-0025197.

XX PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.
 XX N-PSDB; AAZ19368.

PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions

XX Disclosure; Page 205-208; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T. B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
 CC AAY39225 are used in the exemplification of the present invention.

XX SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHVDDIIGTSTWQEAABAVQARDSDVDIRVARVIEDQMAVDSAGKITVRI 60
 DB 1 MGHHHHHVDDIIGTSTWQEAABAVQARDSDVDIRVARVIEDQMAVDSAGKITVRI 60
 QY 61 KLEVSFMRPAQRCQKPPSGPFGAGATVATTPASSPVTLAETGTLPLPLNFWG 120
 DB 61 KLEVSFMRPAQRCQKPPSGPFGAGATVATTPASSPVTLAETGTLPLPLNFWG 119
 QY 121 PAFHERYPNVITTAQTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNIALAISA 180
 DB 120 PAFHERYPNVITTAQTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMNIALAISA 179
 QY 181 QQVYNLPGVSEHLKNGKVLAMYGTTKTWDDPOIALNPGVNLPGTAVVPLHRSOGS 240
 DB 180 QQVYNLPGVSEHLKNGKVLAMYGTTKTWDDPOIALNPGVNLPGTAVVPLHRSOGS 239
 QY 241 GDTFLFTQVLSKQDPGKSGSPFGTGTVDPAVPGALGENGNGMVTCGAETPGCVAYIG 300
 DB 240 GDTFLFTQVLSKQDPGKSGSPFGTGTVDPAVPGALGENGNGMVTCGAETPGCVAYIG 299
 QY 301 ISFLDQASQRLGEALGNSSGNFLPLDQASTQAAAGFASKTPANQALISMIDGAPDGY 360
 DB 300 ISFLDQASQRLGEALGNSSGNFLPLDQASTQAAAGFASKTPANQALISMIDGAPDGY 359
 QY 361 PIINYEYAIYVNNRQKDAATQTLQAFHLWAIIDGNKASFLDQVHFQPLPPAVVKLSDALI 420
 DB 360 PIINYEYAIYVNNRQKDAATQTLQAFHLWAIIDGNKASFLDQVHFQPLPPAVVKLSDALI 419

QY 421 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDOVESTAGSLQGWGAGTAQAQAAVV 480
 DB 420 ATISSAEMKTDAAATLAQEAAGNFERISGDLTKTQIDOVESTAGSLQGWGAGTAQAQAAVV 479
 QY 481 RFOBAANKQKQELDEISTNIRQAGVQYSRADBEQQQALSSQMGFTOSQTVTVVQOQELNR 540
 DB 480 RFOBAANKQKQELDEISTNIRQAGVQYSRADBEQQQALSSQMGFTV--PTTAASPPSTAAA 537
 QY 541 ANEVEAPWADPPDTPVITPCELTAANKAAQOLVLSADNWEYLAAGAKERQRLATSURNA 600
 DB 538 PPAPATPVAPPPPPAAANTPNAQCDPNAPPPADPNAPPPVPIAPNAPQVPR----- 589
 QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGSDS-----SAELTDTPRVATAGPNNF- 653
 DB 590 -----IDNPVGFSFALPAGWVESDAAHFDYGSALLS-----KITGDPPPP 630
 QY 654 -----MDLKEAARKLETGQGLASLAHFADGWNFTNLTQGVKRRGFD 697
 DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
 QY 698 NWEGDAATACEASLDQORQWILHWAKLSAAMAKQAQVVAQLHVWARRHPTVEDIVGLER 757
 DB 676 PYPOTRINQETVSLD-----ANGVSGSASYEVKFSKPNQIWTGVIGSPA 724
 QY 758 LYAENPARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 DB 725 ANAPDAGPPQORWFVW-----LGTANNPVYDKGAALAEISIRPLVAPPPA---P 770
 QY 810 PPPQEQGLIPGLPMPSPDGSVTPGTGMPAAMPVPTGSPGGGLPA 856
 DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQTLPA 802

RESULT 7
 AAY39081

ID AAY39081 standard; Protein; 802 AA.

XX AAY39081;

XX DT 05-NOV-1999 (first entry)

XX M tuberculosis fusion protein Tbf-6.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO9942118-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03265.

XX PR 05-MAY-1998; 98US-0072596.

XX PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.
 XX N-PSDB; AAZ19245.

XX PT New polypeptide comprising antigenic portions of M. tuberculosis

XX Example 10; Page 316-318; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a

CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHHEHHVVDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60
 Db 1 MGHHEHHVVDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60

Qy 61 KLEVSFKRPAQRCGSKPPSGPETGAGACTVATTTPASSPVTLAETGSLLYPLFNLWG 120
 Db 61 KLEVSFKRPAQRCGSKPPSGPETGAGACTVATTTPASSPVTLAETGSLLYPLFNLWG 119

Qy 121 PAFHERYPNVTTTAQGTGSGAGIAQAAGTNYIGASDAYLSEGDMAAHKGLMNTALISA 180
 Db 120 PAFHERYPNVTTTAQGTGSGAGIAQAAGTNYIGASDAYLSEGDMAAHKGLMNTALISA 179

Qy 181 QQVYNLPGVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
 Db 180 QQVYNLPGVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

Qy 241 GDTFLFTQYLSKQDPGKSGPGFTTVDFFPAVFGALGNGNGMVGTCBTPCCVAYIG 300
 Db 240 GDTFLFTQYLSKQDPGKSGPGFTTVDFFPAVFGALGNGNGMVGTCBTPCCVAYIG 299

Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPADGY 360
 Db 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPADGY 359

Qy 361 PIINYEYAVNNRQKDAATQLOAFHWAITDGNKASFQDFVHFQPLPVAWKLSDALI 420
 Db 360 PIINYEYAVNNRQKDAATQLOAFHWAITDGNKASFQDFVHFQPLPVAWKLSDALI 419

Qy 421 ATISSAEMKTDAATLAQAGNFERISGDLTKQIDOVESTAGSLOGWRGAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAATLAQAGNFERISGDLTKQIDOVESTAGSLOGWRGAAGTAAQAAV 479

Qy 481 RFQEAANKQKQELDEISTNIRQAGVQYGRADDEEQQALSSQWGTQSTQTVTVDDQELNR 540
 Db 480 RFQEAANKQKQELDEISTNIRQAGVQYGRADDEEQQALSSQWGFV--PTTAASPPSTAAA 537

Qy 541 ANEVEAPVADPTDVTPTCELTAANKAAQQLVLSADNMREYLAAGAKERORLATSUNA 600
 Db 538 PPAPATPVAPPPAAAANTFNAQPGDPNAAAPPADPNAPPVPIAPNAPQPVK----- 589

Qy 601 AKAYCEVDEEAATALDNDGEGTVQAESAGAVGDS-----SAELTDTPRVATAGEBNF- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPFPF 630

Qy 654 -----MDLKEARKLETGDCQASLAHFAWGNTFNLTLOGDVKRFKRGFD 697
 Db 631 GQPPVPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGCF--YM 675

Qy 698 NWEGDAATACERASLDQOQWILHMAKLSAAMAKQAQYVAQLHVHARREHPYIEIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSPKNGQIWTGVIGSPA 724

Qy 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEFPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPQRFVVM-----LGTANNPVDKGAAKALAESIRPLVAPPPA----P 770

Qy 810 PPPPQEQGLIPGLFMPDSSGSGVTGCTGMPAAAPVPTPGSGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTTQRTLEPA 802

RESULT 8

AAV39033
 ID AAY39033 standard; Protein; 802 AA.
 XX
 AC AAY39033;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbF-2.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W0942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Reto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219156.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 10; Page 251-253; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 XX
 SQ Sequence 802 AA;

Query Match 53.2%; Score 2700; DB 20; Length 802;
 Best Local Similarity 66.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

Qy 1 MGHHEHHVVDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60
 Db 1 MGHHEHHVVDIIGTSPTSWEQAAAEAVQARSDVDIRVARVIEQDMVDSAGKITRYI 60

Qy 61 KLEVSFKRPAQRCGSKPPSGPETGAGACTVATTTPASSPVTLAETGSLLYPLFNLWG 120
 Db 61 KLEVSFKRPAQRCGSKPPSGPETGAGACTVATTTPASSPVTLAETGSLLYPLFNLWG 119

Qy 121 PAFHERYPNVTTTAQGTGSGAGIAQAAGTNYIGASDAYLSEGDMAAHKGLMNTALISA 180
 Db 120 PAFHERYPNVTTTAQGTGSGAGIAQAAGTNYIGASDAYLSEGDMAAHKGLMNTALISA 179

Qy 181 QQVYNLPGVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
 Db 180 QQVYNLPGVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239

Qy 241 GDTFLFTQYLSKQDPGKSGPGFTTVDFFPAVFGALGNGNGMVGTCBTPCCVAYIG 300
 Db 240 GDTFLFTQYLSKQDPGKSGPGFTTVDFFPAVFGALGNGNGMVGTCBTPCCVAYIG 299

Qy 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPADGY 360
 Db 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPADGY 359

QY 361 PIINYEYAI VNNRQKDAATQATLQAFHLWAI TDGNKASFLDQVHFQPLPPAVVKS DALI 420
 Db 360 PIINYEYAI VNNRQKDAATQATLQAFHLWAI TDGNKASFLDQVHFQPLPPAVVKS DALI 419
 QY 421 ATISSAEMKTDAA TLQAEAGNFERISGDLKTQIDQVESTAGSIQOGWRGAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAA TLQAEAGNFERISGDLKTQIDQVESTAGSIQOGWRGAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTQSVTVDDQBELNR 540
 Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTV--PTTAAASPPSTAAA 537
 QY 541 ANEVEAPMADPPTDVTTPCELTAANKAAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATPVAPPPAAANTPNAQGDNPAPPPADPNAPPPVIAFNAPOFVR-----589
 QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SAELTDTPRVATAGEPNF- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPPPP 630
 QY 654 -----MDLKEARKLETGDOGLASLAHFAFGWNTFNLTQGVKFRGPD 697
 Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675
 QY 698 NWEGDAATACEASLDQOROWILHMAKLSAAMAKQAQYVAQLHWARRREFTYEDIVGLER 757
 Db 676 PYFGTRINQETVSLD-----ANGVSGSASYEVKFSDPKPNQIWTGVIGSPA 724
 QY 758 LYAENPSARDQILPVVAEYQORSEKVLTEYNN-----KAALFPVNP-PKPPPAIKIDP 809
 Db 725 ANAPDAGPPORWFWVW-----LGTANNPVDKGAALAESIRPLVAPPPA----P 770
 QY 810 PPPPQOGLIPGLFMPDPSGSGVTGCTGMPAPMPVPTGSPGCGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTTPORTLPA 802

RESULT 9

ID AAU74592 standard; Protein; 802 AA.
 AC AAU74592;
 XX 08-MAY-2002 (first entry)
 DT XX
 DE Antigenic fusion protein Tbra3-38kd-Tb38-1-DPEP (TbF-2).
 DE XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; Tbra3-38kd-Tb38-1-DPEP; TbF-2.
 XX
 OS Chimeric - Mycobacterium tuberculosis.
 XX
 XX US2002009459-A1.
 XX
 PD 24-JAN-2002.
 XX
 PF 07-APR-1999; 99US-0287849.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 01-OCT-1997; 97US-0942578.
 PR 18-FEB-1998; 98US-0025197.
 PR 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 XX
 PA (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX WPI; 2002-171134/22.
 DR

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis -
 XX Claim 1; Fig 5G-J; 62pp; English.
 XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention.
 CC Note: The specification states that this polypeptide is encoded by the
 CC polynucleotide shown in ABK14132.
 XX Sequence 802 AA;
 SQ
 Query Match 53.2%; Score 2700; DB 23; Length 802;
 Best Local Similarity 56.6%; Pred. No. 2.9e-158;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
 QY 1 MGHHHHHVHDIIGTSPTEQAAAEAVQEARSDVDIRVARVIEQDMVDSAGKITRYI 60
 Db 1 MGHHHHHVHDIIGTSPTEQAAAEAVQEARSDVDIRVARVIEQDMVDSAGKITRYI 60
 QY 61 KLEVSFRRPAQRCSKPSGSPETGAGAGTATTTPASSPVTLAETGTLTYLFLNMG 120
 Db 61 KLEVSFRRPAQR--GSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTYLFLNMG 119
 QY 121 PAFHERYPNTITTAQGTSGAGIAQAAAGTNTGASDAYLSEGDMAAHKGLMIALAISA 180
 Db 120 PAFHERYPNTITTAQGTSGAGIAQAAAGTNTGASDAYLSEGDMAAHKGLMIALAISA 179
 QY 181 QQVNNYLPVGVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 240
 Db 180 QQVNNYLPVGVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 239
 QY 241 GDTFLFTQYLSKQDPEGWKSQPGFTTVDPPAVPGALGENGNGMTGCAETGCVAYIG 300
 Db 240 GDTFLFTQYLSKQDPEGWKSQPGFTTVDPPAVPGALGENGNGMTGCAETGCVAYIG 299
 QY 301 ISFLDQASQKGLGEAQLGNSNGNLLPDAQSICAAAAGFASKTPANCAISMIDGPADGY 360
 Db 300 ISFLDQASQKGLGEAQLGNSNGNLLPDAQSICAAAAGFASKTPANCAISMIDGPADGY 359
 QY 361 PIINYEYAI VNNRQKDAATQATLQAFHLWAI TDGNKASFLDQVHFQPLPPAVVKS DALI 420
 Db 360 PIINYEYAI VNNRQKDAATQATLQAFHLWAI TDGNKASFLDQVHFQPLPPAVVKS DALI 419
 QY 421 ATISSAEMKTDAA TLQAEAGNFERISGDLKTQIDQVESTAGSIQOGWRGAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAA TLQAEAGNFERISGDLKTQIDQVESTAGSIQOGWRGAAGTAAQAAV 479
 QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTQSVTVDDQBELNR 540
 Db 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADEBQQQALSSQMGFTV--PTTAAASPPSTAAA 537
 QY 541 ANEVEAPMADPPTDVTTPCELTAANKAAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATPVAPPPAAANTPNAQGDNPAPPPADPNAPPPVIAFNAPOFVR-----589
 QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SAELTDTPRVATAGEPNF- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPPPP 630
 QY 654 -----MDLKEARKLETGDOGLASLAHFAFGWNTFNLTQGVKFRGPD 697
 Db 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFE--YM 675

```

QY 698 NWEGDAATACASLDQQRQWILHMAKLSAAVAKAQYVAQLHVWARREHPTIEDIVGLER 757
D 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKNGQIWTGVIQSPA 724
QY 758 LYAENPSARDQILPVYAAYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
D 725 ANAPDAGPPQRFVFW-----IGTANNPVVDKGAAKALAESIRPLVAPPPA-----P 770
QY 810 PPPPOEQGLIPGFLMPPSDGSGVTPGTGMPAAPVPPPTGSGGGGLPA 856
D 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTFORTLPA 802

RESULT 10
ID AAW81737
XX
AC AAW81737;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide TbH4-Xp1 #1.
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
FN WO9816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18293.
XX
PR 13-MAR-1997; 97US-0818112.
PR 11-OCT-1996; 96US-0730510.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-261042/23.
DR N-PSDB; AAV64549.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and for diagnosis, treatment and prevention of tuberculosis
XX
PS Example 3c; Page 178-179; 230pp; English.
XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis.
XX
SQ Sequence 460 AA;
Query Watch 46.8%; Score 2375; DB 19; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.6e-138;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 525 TQSQTVTVDQOELNRAVEAPMADPFDVITPCELTAANKAAQQLVLSADNNREYLA 584
D 2 TQSQTVTVDQOELNRAVEAPMADPFDVITPCELTAANKAAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEATALNDGEGTVQAESAGAVGDSAEITDTPR 644
D 62 AGAKERQRLATSLRNAKAYGEVDEEATALNDGEGTVQAESAGAVGDSAEITDTPR 121

```

```

QY 645 VATAGEPNFMDLKEAARKLETGQGASLAHPADGWNFTNLTOGDKVRFGRFDNWEGDAA 704
D 122 VATAGEPNFMDLKEAARKLETGQGASLAHPADGWNFTNLTOGDKVRFGRFDNWEGDAA 181
QY 705 TACEASLDQOROWILHMAKLSAAMAKAQYVAQLHVWARREHPTIEDIVGLERLYAENPS 764
D 182 TACEASLDQOROWILHMAKLSAAMAKAQYVAQLHVWARREHPTIEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAAYQORSEKVLTEYNNKAALEFVNPPKPPPAIKIDPPPPPOEQGLIPGFLM 824
D 242 ARDQILPVYAAYQORSEKVLTEYNNKAALEFVNPPKPPPAIKIDPPPPPOEQGLIPGFLM 301
QY 825 PPSDGSVTPGTGMPAAPVPPPTGSGGGGLPADTAAQLTSAGREAAALSGDVAVKAASLG 884
D 302 PPSDGSVTPGTGMPAAPVPPPTGSGGGGLPADTAAQLTSAGREAAALSGDVAVKAASLG 361
QY 885 GGGGGVPSAPLGSAGIAESVAPAGAGDIAGLQGRAGGAGALGGGGMGMPMAAHQOQ 944
D 362 GGGGGVPSAPLGSAGIAESVAPAGAGDIAGLQGRAGGAGALGGGGMGMPMAAHQOQ 421
QY 945 GGAKSQSGQDEALYTEDRANTEAVIGNRRRODSKESK 983
D 422 GGAKSQSGQDEALYTEDRANTEAVIGNRRRODSKESK 460

RESULT 11
ID AAW64367
XX
AC AAW64367;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-4-Xp1.
XX
KW Tuberculosis; infection; diagnosis; antigen; TbH-4; Xp1.
OS Mycobacterium tuberculosis strain H37RV.
OS Mycobacterium tuberculosis strain Erdman.
XX
FN WO9816645-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US18214.
PR 13-MAR-1997; 97US-0818111.
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1998-251292/22.
DR N-PSDB; AAV44440.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Example 3; Page 189-191; 250pp; English.
XX
CC This polypeptide is the predicted amino acid sequence of
CC Mycobacterium tuberculosis antigen TbH4-Xp1. The sequence was
CC deduced from a composite sequence (see AAV44440) of overlapping DNA
CC clones TbH4 and Xp1. TbH4 (see also AAW64319) was isolated from a
CC M. tuberculosis strain H37Rv expression library and Xp1 from a M.
CC tuberculosis Erdman expression library. Recombinant Xp1 protein
CC was prepared. It stimulates cell proliferation and interferon-
CC gamma production in T cells isolated from M. tuberculosis-immune
CC donors. The invention relates to methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAW64291-W64379)

```


KW vaccine; immunity.
XX Mycobacterium tuberculosis.
XX WO9942118-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03265.
XX 05-MAY-1998; 98US-0072596.
XX 18-FEB-1998; 98US-0024753.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Iwardzik DR, Vedvick TS;
XX WPI; 1999-527416/44.
XX N-PSDB; AA219138.
XX New polypeptide comprising antigenic portions of M. tuberculosis
PT Example 3; Page 219-220; 33pp; English.
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX
SQ Sequence 460 AA;
Query Match 46.8%; Score 2375; DB 20; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.6e-138; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 0;
QY 525 TQSQVTVDQQLIRANEVEAPMDPTDVTITPCELTAARNAQAQQLVLSADNNREYLA 584
Db 2 TQSQVTVDQQLIRANEVEAPMDPTDVTITPCELTAARNAQAQQLVLSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGEVDEEATATLNDGEGTVQAESAGVGDSSAELTDTPR 644
Db 62 AGAKERQRLATSLRNAKAYGEVDEEATATLNDGEGTVQAESAGVGDSSAELTDTPR 121
QY 645 VATAGEPNFMDLKEARKLETGQASLAHFADGNTNLTLOQGVKFRGFDNWEGBAA 704
Db 122 VATAGEPNFMDLKEARKLETGQASLAHFADGNTNLTLOQGVKFRGFDNWEGBAA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKQAYVAQLHWARREHTYEDIVGLERLYAENPS 764
Db 182 TACEASLDQORQWILHMAKLSAAMAKQAYVAQLHWARREHTYEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPBPPPAIKIDPPPPQGLIPGLM 824
Db 242 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPBPPPAIKIDPPPPQGLIPGLM 301
QY 825 PPSDGSVTPGTGMPAAPVPTGSPGGGLPADTAQLTSAAGRAAALSGLDVAVKAASLG 884
Db 302 PPSDGSVTPGTGMPAAPVPTGSPGGGLPADTAQLTSAAGRAAALSGLDVAVKAASLG 361
QY 885 GGGGGVPSAPLGSAGAESVRPAGAGDIAGLGGGRAGGGAALGGGGMGPMGAHQOQ 944
Db 362 GGGGGVPSAPLGSAGAESVRPAGAGDIAGLGGGRAGGGAALGGGGMGPMGAHQOQ 421
QY 945 GGAKSQSGQDEEALYTEDRAWTEAVIGNRRRQDSKESK 983
Db 422 GGAKSQSGQDEEALYTEDRAWTEAVIGNRRRQDSKESK 460

RESULT 14

AA39225
ID AAY39225 standard; Protein; 652 AA.
XX AAY39225;
AC AAY39225;
XX 05-NOV-1999 (first entry)
XX M. tuberculosis fusion protein Tbf-8 amino acid sequence.
DE
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
XX WO9942076-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03268.
XX 05-MAY-1998; 98US-0072967.
XX 18-FEB-1998; 98US-0025197.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Iwardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.
XX N-PSDB; AA219460.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX
XX Claim 37; Page 274-276; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 652 AA;
Query Match 36.3%; Score 1839; DB 20; Length 652;
Best Local Similarity 48.5%; Pred. No. 3.2e-105;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHVHDIIGTSTPSWEQAAAEAVQARSDVDIRVIEQDMAYDSAGKITRYI 60
Db 1 MGHHHHHH----- 8
QY 61 KLEVSFKVRPAQPRCGSKPPSPETGAGAGTVATTTPASSPVTIAETGTLTYPLFNLWG 120
Db 9 -----GSKPFGSGPETGAGAGTVATTTPASSPVTIAETGTLTYPLFNLWG 53
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNTGASDAYLSEGDMAAHKLMMIALAISA 180
Db 54 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNTGASDAYLSEGDMAAHKLMMIALAISA 113
QY 181 QQVNNYLPVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 114 QQVNNYLPVSEHLKNGKVLAAVYOGTITKTDWDPQIAALNPGVNLPGTAVVPLHRS DGS 173
QY 241 GDTFLFTQYLSKQDPGKMGKSPQGTGTTVDPAVPALGNGNGMGTGCAETPCGVAYIG 300

Db 174 GDTFLFTQYLSKQDPEGKSPGFGTTVDPPAVFGALGENGGMTGCAETPGCVAYIG 233
Qy 301 ISFLDQASQRLGEAQLGNSGNFLPDQAOSIQAAAAGFASKTPANQAISMIDGPAPDGY 360
Db 234 ISFLDQASQRLGEAQLGNSGNFLPDQAOSIQAAAAGFASKTPANQAISMIDGPAPDGY 293
Qy 361 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVKS DALI 420
Db 294 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVKS DALI 353
Qy 421 ATISSAEMKTDATLAQEAQNFERISGDLKTQIDQVESTAGSLQGWRAAGTAQAAYV 480
Db 354 ATISSGG-----GSGGGSGGSGSVPTTAASPSTA-----AAPAPATPVA 396
Qy 481 RFQEAANKQKQELDEISTNIRQAGVQVSRADDEQQQALSSQMGTQSTQVTVDDQOEILNR 540
Db 397 PPPPAA-----ANTPNAQFG----- 411
Qy 541 ANEVEAPMADPPTDVPITPCELTAAKVAQAQQLVLSADN-----MREYLAAGAKERQ----- 591
Db 412 -----DENAAPPADPNAPPVPIAPNAPQPVRI--DNVPGGFSFALPAGWVESDAAHFD 464
Qy 592 -----RLATSLRNAAKAYGEVDEEAATALDND-- 618
Db 465 YGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDQKLYASAEA---TDSKAARLGSDMG 521
Qy 619 -----GGTVOAESAGAVGSDSAELTDTTPRVATAGEPNFMDLKEAARKLETGD 667
Db 522 EFYMPYPGTRINQETVSLDANGVSGSASYEV-----KFSDESKPNGQIWTGV 569
Qy 668 QGASLAHAFDQWTFNLTLOGDVKFRGPDNWDGDAATACASLDQORQWILMAKLSAA 727
Db 570 IGSPAANAPDA-----GPPQRM--FVVMILGTANNPVDK----- 600
Qy 728 MAKQAQYVAQLHVHVARREHPTYEDIVGLERLYAENPSARDQILPVVAYEQORSEKVLTEY 787
Db 601 -----GAAKALAE--SIRPLVAP----- 616
Qy 788 NNKAALFVNNPPKPPPAIKIDPPPPQBOGLIPGFLMPPSDSGSVTPGTCMPAAPWVPT 847
Db 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPTT 643
Qy 848 GSPGGGLPA 856
Db 644 PTQRTILPA 652

RESULT 15

AAV39082
ID AAV39082 standard; Protein; 652 AA.

AC AAV39082;

XX 05-NOV-1999 (first entry)

DE M tuberculosis fusion protein Tbf-8.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO9942118-A2.

PD 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527416/44.
DR N-PSDB; AAZ19248.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis
PS Example 10; Page 320-321; 323pp; English.
XX
CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

XX SQ Sequence 652 AA;

Query Match 36.3%; Score 1839; DB 20; Length 652;

Best Local Similarity 48.5%; Pred. No. 3.2e-105;

Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;

Qy 1 MGHHHHHHVDIIIGTSPTSWEQAAAQVARDSDVDIRVARVIEQDMAVDSAGKITRYI 60

Db 1 MGHHHHHH----- 8

Qy 61 KLEVSFXWRPAQPCRGKPPSGPETGAGAGTATTTPASSPVTLAETGSLLYPLFNLWG 120

Db 9 -----GSKPFGSGPETGAGAGTATTTPASSPVTLAETGSLLYPLFNLWG 53

Qy 121 PAFHERYPNVITTAQGTGSGAGIAQAAAAGTNNIGASDAYLSEGDMAHKLMMIALAISA 180

Db 54 PAFHERYPNVITTAQGTGSGAGIAQAAAAGTNNIGASDAYLSEGDMAHKLMMIALAISA 113

Qy 181 QOVNYYNPGVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 240

Db 114 QOVNYYNPGVSEHLKNGKVLAAVYQGTIKTDDPQIAALNPGVNLPGTAVVPLHSDGS 173

Qy 241 GDTFLFTQYLSKQDPEGKSPGFGTTVDPPAVFGALGENGGMTGCAETPGCVAYIG 300

Db 174 GDTFLFTQYLSKQDPEGKSPGFGTTVDPPAVFGALGENGGMTGCAETPGCVAYIG 233

Qy 301 ISFLDQASQRLGEAQLGNSGNFLPDQAOSIQAAAAGFASKTPANQAISMIDGPAPDGY 360

Db 234 ISFLDQASQRLGEAQLGNSGNFLPDQAOSIQAAAAGFASKTPANQAISMIDGPAPDGY 293

Qy 361 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVKS DALI 420

Db 294 PIINYEYAI VNNRQKDAATQTLQAFLHWAITDGNKASFLDQVHFQPLPPAVVKS DALI 353

Qy 421 ATISSAEMKTDATLAQEAQNFERISGDLKTQIDQVESTAGSLQGWRAAGTAQAAYV 480

Db 354 ATISSGG-----GSGGGSGGSGSVPTTAASPSTA-----AAPAPATPVA 396

Qy 481 RFQEAANKQKQELDEISTNIRQAGVQVSRADDEQQQALSSQMGTQSTQVTVDDQOEILNR 540

Db 397 PPPPAA-----ANTPNAQFG----- 411

Qy 541 ANEVEAPMADPPTDVPITPCELTAAKVAQAQQLVLSADN-----MREYLAAGAKERQ----- 591

Db 412 -----DENAAPPADPNAPPVPIAPNAPQPVRI--DNVPGGFSFALPAGWVESDAAHFD 464

Qy 592 -----RLATSLRNAAKAYGEVDEEAATALDND-- 618

Db 465 YGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDQKLYASAEA---TDSKAARLGSDMG 521

Qy 619 -----GGTVOAESAGAVGSDSAELTDTTPRVATAGEPNFMDLKEAARKLETGD 667

Db 522 EFYMPYPGTRINQETVSLDANGVSGSASYEV-----KFSDESKPNGQIWTGV 569

```

QY 668 QGASLAHPADGNTFNLTLOQDVKFRGFDNWEGDAATACEASLDQOROWILHMAKLSAA 727
Db 570 IGSPAANAPDA-----GPPQKW--FVVMGTANNPVDK----- 600
QY 728 MAKQAOYVAQLHVWARREHPTYEIVGLERLYAENPSARDQILPVYAEYQORSEKVLTEY 787
Db 601 -----GAKALAE--SIRPLVAP----- 616
QY 788 NKAALPEVNPCKPPPAIKIDPPPPPPQOGLIPGFLMPPSDGSGVTGTMGMPAAAPMVPT 847
Db 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPTT 643
QY 848 GSPGGGLPA 856
Db 644 PTPQRTLEA 652

```

Search completed: November 21, 2003, 16:03:20
Job time : 63.515 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 20.7299 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHVHHVHVDIIGTSPTSW.....RANWEAVIGNRRRODSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pap:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pap:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pap:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pap:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pap:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2700	53.2	802	4	US-09-056-556-214
2	2700	53.2	802	4	US-09-072-596-209
3	2700	53.2	802	4	US-09-072-596-346
4	2375	46.8	460	4	US-09-056-556-184
5	2375	46.8	460	4	US-09-072-596-179
6	1839	36.3	652	4	US-09-072-596-350
7	1831	36.1	374	3	US-08-818-112-153
8	1831	36.1	374	4	US-08-818-111-118
9	1831	36.1	374	4	US-09-056-556-153
10	1831	36.1	374	4	US-09-056-556-155
11	1831	36.1	374	4	US-09-072-596-148
12	1831	36.1	374	4	US-09-072-596-150
13	1829	36.1	373	4	US-09-118-426-5
14	1820	35.9	351	4	US-09-118-426-6
15	791	15.6	166	3	US-08-818-112-89
16	791	15.6	166	4	US-08-818-111-90
17	791	15.6	166	4	US-09-056-556-89
18	791	15.6	166	4	US-09-072-596-90
19	485	9.6	100	3	US-08-818-112-115
20	485	9.6	100	4	US-08-818-111-110
21	485	9.6	100	4	US-09-056-556-115
22	485	9.6	100	4	US-09-072-596-110
23	481	9.5	100	4	US-09-116-492A-5
24	466	9.2	95	3	US-08-818-112-88
25	466	9.2	95	4	US-08-818-111-89
26	466	9.2	95	4	US-09-056-556-88
27	466	9.2	95	4	US-09-072-596-89

Sequence 2, Appli
Sequence 2, Appli
Sequence 117, App
Sequence 117, App
Sequence 112, App
Sequence 112, App
Sequence 112, App
Sequence 251, App
Sequence 379, App
Sequence 73, Appl
Sequence 74, Appl
Sequence 73, Appl
Sequence 74, Appl
Sequence 77, Appl
Sequence 78, Appl
Sequence 78, Appl
Sequence 6, Appli
Sequence 4953, Ap

28 413 8.1 348 1 US-07-989-845-2
29 413 8.1 348 5 PCT-US93-11298-2
30 392 7.7 80 3 US-08-818-112-117
31 392 7.7 80 4 US-08-818-111-112
32 392 7.7 80 4 US-09-056-556-117
33 392 7.7 80 4 US-09-072-596-112
34 378 7.5 432 4 US-08-311-731A-251
35 335.5 6.6 372 4 US-08-311-731A-379
36 329.5 6.5 364 3 US-08-818-112-73
37 329.5 6.5 364 4 US-08-818-111-74
38 329.5 6.5 364 4 US-09-056-556-73
39 329.5 6.5 364 4 US-09-072-596-74
40 324 6.4 66 3 US-08-818-112-77
41 324 6.4 66 4 US-08-818-111-78
42 324 6.4 66 4 US-09-056-556-77
43 324 6.4 66 4 US-09-072-596-78
44 236 4.7 49 4 US-09-116-492A-6
45 226.5 4.5 301 4 US-09-107-532A-4953

ALIGNMENTS

RESULT 1
US-09-056-556-214
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-214

Query Match 53.2%; Score 2700; DB 4; Length 802;
Best Local Similarity 66.6%; Pred. No. 5e-187;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
Oy 1 MGHVHHVHVDIIGTSPTSWEQAAAVQARSDVDIRVARVIEQDMVDSAGKITYRI 60
Db 1 MGHVHHVHVDIIGTSPTSWEQAAAVQARSDVDIRVARVIEQDMVDSAGKITYRI 60
Oy 61 KLEVSFKMRPAQPCGSKPSPGSGPETGAGAGTATTATGAGTLLPLFLNLWG 120

Db 61 KLEVSFVRPAQPR-GSKPSPGSGPETGAGAGTATTTPASSPVTLAETGSLTYPLFNLWG 119
QY 121 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 179
QY 181 QOVNYPVSEHLKNGKVLAAVYOGTITKTDPOIAALNPGVNLPGTAVPLHRS DGS 240
Db 180 QOVNYPVSEHLKNGKVLAAVYOGTITKTDPOIAALNPGVNLPGTAVPLHRS DGS 239
QY 241 GDTFLFTQYLSKODPEGWKSPPGFTTVDPPFVPGALGENGGMTGCAETPCGVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKSPPGFTTVDPPFVPGALGENGGMTGCAETPCGVAYIG 299
QY 301 ISFLDQASQRLGEAQLGNSGNFLPLDPAQSIQAAAAGFASKTPANQAI SMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSGNFLPLDPAQSIQAAAAGFASKTPANQAI SMIDGPADGY 359
QY 361 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVAVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVAVKLS DALI 419
QY 421 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQAAV 480
Db 420 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQAAV 479
QY 481 RFQEAANKQKQELDEISTNIRQAGVQYSRADSEQQQALSSQMGFV--PTTAA SPSTAAA 540
Db 480 RFQEAANKQKQELDEISTNIRQAGVQYSRADSEQQQALSSQMGFV--PTTAA SPSTAAA 537
QY 541 ANEVEAPMADPTDVTPTPCELTAANKAAQVLSDNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATVAPPPAAAANTPNAQCDNAPPPADPNAPPPVPIAPNAPQVR----- 589
QY 601 AKAYGEVDEEATALNDGSGTVOASAGAVGDS-----SALFDTT PRVTAGEPNF- 653
Db 590 -----IDNPVGGFSFALPAGWVESDAAHFDYGSALLS-----KTTGDP PPP 630
QY 654 -----MDLKEARAKLETGQGLASLAHFDGWNFTNLTLOGGVKBERGPD 697
Db 631 GQPPPVANDTRIVLGRJDDQKLYASAEATDSKAAA-----RLGSDMGBF--YM 675
QY 698 NWEGDAATACEASLDQORQWILHMAKLSAAMAKQAQVLAQLHWARREHPTVEDIVGLER 757
Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDPKPNQIWTGVIGSPA 724
QY 758 LYAENPSARDQIILPVYAQYQORSEKVLTEYNN-----KALEPVNP-PKPPPAIKIDP 809
Db 725 ANAPDAGPPORFVFW-----LGTANNPVDKGAALAESTRPLVAPPPA-----P 770
QY 810 PPPPQEQGLIPGLMPSPDGSVTPGTGMPAPMPVPTGSPGGGLPA 856
Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTQRTLPA 802

RESULT 2

US-09-072-596-209
; Sequence 209, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/072,596
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-209
Query Match 53.2%; Score 2700; DB 4; Length 802;
Best Local Similarity 66.68; Pred. No. 5e-187;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHHHHHHVIDIIGTSPSWEQAAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYI 60
Db 1 MGHHHHHHVIDIIGTSPSWEQAAAAEAVQARSDVDDIRVARVIEQDMAVDSAGKITRYI 60
QY 61 KLEVSFVRPAQPRGSKPSPGSGPETGAGAGTATTTPASSPVTLAETGSLTYPLFNLWG 120
Db 61 KLEVSFVRPAQPRGSKPSPGSGPETGAGAGTATTTPASSPVTLAETGSLTYPLFNLWG 119
QY 121 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 180
Db 120 PAFHERYPNVTITTAQGTSGAGTAQAAAGTNNIGASDAYLSEGDMAHKLGMNIALAISA 179
QY 181 QOVNYPVSEHLKNGKVLAAVYOGTITKTDPOIAALNPGVNLPGTAVPLHRS DGS 240
Db 180 QOVNYPVSEHLKNGKVLAAVYOGTITKTDPOIAALNPGVNLPGTAVPLHRS DGS 239
QY 241 GDTFLFTQYLSKODPEGWKSPPGFTTVDPPFVPGALGENGGMTGCAETPCGVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKSPPGFTTVDPPFVPGALGENGGMTGCAETPCGVAYIG 299
QY 301 ISFLDQASQRLGEAQLGNSGNFLPLDPAQSIQAAAAGFASKTPANQAI SMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSGNFLPLDPAQSIQAAAAGFASKTPANQAI SMIDGPADGY 359
QY 361 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVAVKLS DALI 420
Db 360 PIINYEYAI VNNRQKDAATQATQAFLEHWAITDGNKASFLDQVHFQPLPVAVKLS DALI 419
QY 421 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQAAV 480
Db 420 ATISSAEMKTDATLAQEAQNFERRISGDLKTQIDQVESTAGSLQGGWRGAAGTAAQAAV 479
QY 481 RFQEAANKQKQELDEISTNIRQAGVQYSRADSEQQQALSSQMGFV--PTTAA SPSTAAA 540
Db 480 RFQEAANKQKQELDEISTNIRQAGVQYSRADSEQQQALSSQMGFV--PTTAA SPSTAAA 537
QY 541 ANEVEAPMADPTDVTPTPCELTAANKAAQVLSDNMREYLAAGAKERQRLATSLRNA 600
Db 538 PPAPATVAPPPAAAANTPNAQCDNAPPPADPNAPPPVPIAPNAPQVR----- 589

QY 601 AKAYGEVDEEATAALDNDGEGTVOAESAGAVGGDS-----SABLTDTPRVATAGEPNF- 653
DB 590 -----IDNPVGGFSFALPAGWVESAAAHFDYGSALLS-----KITGDPFPF 630
QY 654 -----MDLKEARKLETGDCASLAHFAFGWNTFNLTLQGDVKKRFRGFD 697
DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEGDAATACEASLDQOROWILHMAKLSAAMAKQAOYVAQLHWARREHPTYEDIVGLER 757
DB 676 PYFGTRINQETVSLD-----ANGVSGSASYVEVKFSDPSKPGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAEIRPLVAPPPA-----P 770
QY 810 PPPPQOGLIPGLMPPSDGSGVTPGTGMPAAPMVPPTGSPGGGLPA 856
DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 3

US-09-072-596-346
; Sequence 346, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-346

Query Match 53.2%; Score 2700; DB 4; Length 802;
Best Local Similarity 66.6%; Pred. No. 5e-187;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHHVIDIGTSPTSWBOAAAABAVORARDSVDDIRVARVIBODMAVDGAGKITVYRI 60
DB 1 MGHHHHHHVIDIGTSPTSWBOAAAABAVORARDSVDDIRVARVIBODMAVDGAGKITVYRI 60
QY 61 KLEVSFKMPAPRCOSKPPSGSPETGAGAGVATTTPASSPVTLAETGTLTLYPLFNWLG 120
DB 61 KLEVSFKMPAPRCOSKPPSGSPETGAGAGVATTTPASSPVTLAETGTLTLYPLFNWLG 119
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTINIGASDAYLSEGDMAAHKGLNIALAISA 180
DB 120 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTINIGASDAYLSEGDMAAHKGLNIALAISA 179
QY 181 QQYNYNLPVGSVSHLKLNGVLAAMYOGTITKTDWDDQIARALNPGVNLPGTAVVPLHRS DGS 240
DB 180 QQYNYNLPVGSVSHLKLNGVLAAMYOGTITKTDWDDQIARALNPGVNLPGTAVVPLHRS DGS 239
QY 241 GDTFLEFQYLSKQDPEGWGKSPFGTITVDFPAPVPGALGENGNGMVTGCAETPGCVAYIG 300
DB 240 GDTFLEFQYLSKQDPEGWGKSPFGTITVDFPAPVPGALGENGNGMVTGCAETPGCVAYIG 299
QY 301 ISFLDOASORGLEAQLGNSGNFLLPDAQSIOAAAANGFASKTPANCAISMIDGRAPDGY 360
DB 300 ISFLDOASORGLEAQLGNSGNFLLPDAQSIOAAAANGFASKTPANCAISMIDGRAPDGY 359
QY 361 PIINYEYAIVNNRQKDAATAQTLQAFELHWAITDGNKASFELDQVHFQPLPPAVVVKLSDALI 420
DB 360 PIINYEYAIVNNRQKDAATAQTLQAFELHWAITDGNKASFELDQVHFQPLPPAVVVKLSDALI 419
QY 421 ATISSAEMKTDATLAQEAQNTFERISGDLKTQIDQVESTAGSLQGWGAAAGTAAQAAV 480
DB 420 ATISSAEMKTDATLAQEAQNTFERISGDLKTQIDQVESTAGSLQGWGAAAGTAAQAAV 479
QY 481 RFOEANKOKOELDEISTNIRQAGVOYSRADDEEQOALSQMGFTQSVTVVDQOEILNR 540
DB 480 RFOEANKOKOELDEISTNIRQAGVOYSRADDEEQOALSQMGFTQSVTVVDQOEILNR 537
QY 541 ANEVEAPMADPTDPTTCELTAANKAAQQLVLSADNMREYLAAGAKERQRLATSLRNA 600
DB 538 PPAPATPVAPPPPPAAAANTENAPGDENAPPPADPNAPPPVPIAPNAPQVPR----- 589
QY 601 AKAYGEVDEEATAALDNDGEGTVOAESAGAVGGDS-----SABLTDTPRVATAGEPNF- 653
DB 590 -----IDNPVGGFSFALPAGWVESAAAHFDYGSALLS-----KITGDPFPF 630
QY 654 -----MDLKEARKLETGDCASLAHFAFGWNTFNLTLQGDVKKRFRGFD 697
DB 631 GQPPPVANDTRIVLGRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
QY 698 NWEGDAATACEASLDQOROWILHMAKLSAAMAKQAOYVAQLHWARREHPTYEDIVGLER 757
DB 676 PYFGTRINQETVSLD-----ANGVSGSASYVEVKFSDPSKPGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQWFFVW-----LGTANNPVDKGAALAEIRPLVAPPPA-----P 770
QY 810 PPPPQOGLIPGLMPPSDGSGVTPGTGMPAAPMVPPTGSPGGGLPA 856
DB 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPQRTLPA 802

RESULT 4

US-09-056-556-184
; Sequence 184, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 210121.457
REFERENCE/DOCKET NUMBER:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-184

Query Match 46.8%; Score 2375; DB 4; Length 460;

Best Local Similarity 100.0%; Pred. No. 7.3e-164; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 0;

```
QY 525 TQSTVTVDQOEILNRANEVEAPMDPTDVTTCPELTAAKNAQAQVLVSADNNREYLA 584
Db 2 TQSTVTVDQOEILNRANEVEAPMDPTDVTTCPELTAAKNAQAQVLVSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTPR 644
Db 62 AGAKERQRLATSLRNAAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTPR 121
QY 645 VATAGEPNFMDLKEAARKLETGQGSALAHFADGWNFTNLTQGDVKRFRGFDNNEGDAA 704
Db 122 VATAGEPNFMDLKEAARKLETGQGSALAHFADGWNFTNLTQGDVKRFRGFDNNEGDAA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 764
Db 182 TACEASLDQORQWILHMAKLSAAMAKAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPPAIKIDPPPPQEOGLIPGFLM 824
Db 242 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPPAIKIDPPPPQEOGLIPGFLM 301
QY 825 PPSDGSVPTGTGMPAAPVPTGSGGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 884
Db 302 PPSDGSVPTGTGMPAAPVPTGSGGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 361
QY 885 GGGGGGVPSAPLGSALGGAESVTPAGAGDIAGLGGGRAGGGAALGGGGMWGAHQQG 944
Db 362 GGGGGGVPSAPLGSALGGAESVTPAGAGDIAGLGGGRAGGGAALGGGGMWGAHQQG 421
QY 945 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 983
Db 422 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 460
```

RESULT 5

US-09-072-596-179
; Sequence 179, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4300
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 179:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-179

Query Match 46.8%; Score 2375; DB 4; Length 460;

Best Local Similarity 100.0%; Pred. No. 7.3e-164;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 525 TQSTVTVDQOEILNRANEVEAPMDPTDVTTCPELTAAKNAQAQVLVSADNNREYLA 584
Db 2 TQSTVTVDQOEILNRANEVEAPMDPTDVTTCPELTAAKNAQAQVLVSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTPR 644
Db 62 AGAKERQRLATSLRNAAKAYGEVDEEAATALDNDGEGTVOAESAGAVGGDSSAELTDTPR 121
QY 645 VATAGEPNFMDLKEAARKLETGQGSALAHFADGWNFTNLTQGDVKRFRGFDNNEGDAA 704
Db 122 VATAGEPNFMDLKEAARKLETGQGSALAHFADGWNFTNLTQGDVKRFRGFDNNEGDAA 181
QY 705 TACEASLDQORQWILHMAKLSAAMAKAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 764
Db 182 TACEASLDQORQWILHMAKLSAAMAKAQYVAQLHWARREHPTTYEDIVGLERLYAENPS 241
QY 765 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPPAIKIDPPPPQEOGLIPGFLM 824
Db 242 ARDQILPVYAEYQORSEKVLTEYNNKAALPEVNPFPKPPAIKIDPPPPQEOGLIPGFLM 301
QY 825 PPSDGSVPTGTGMPAAPVPTGSGGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 884
Db 302 PPSDGSVPTGTGMPAAPVPTGSGGGGLPADTAQAQLTSAGREAAALSGDVAVKAAISG 361
QY 885 GGGGGGVPSAPLGSALGGAESVTPAGAGDIAGLGGGRAGGGAALGGGGMWGAHQQG 944
Db 362 GGGGGGVPSAPLGSALGGAESVTPAGAGDIAGLGGGRAGGGAALGGGGMWGAHQQG 421
```


QY 945 GGAAGKSGQDEALYTEDRAWTEAVICNRRQDSKSK 983
DB 422 GGAAGKSGQDEALYTEDRAWTEAVICNRRQDSKSK 460

RESULT 6

US-09-072-596-350
Sequence 350, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 350:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-072-596-350

Query Match 36.3%; Score 1839; DB 4; Length 652;
Best Local Similarity 48.5%; Pred. No. 7, 7e-125;
Matches 441; Conservative 26; Mismatches 133; Indels 310; Gaps 19;

QY 1 MGHHHHHVHDIIGTPTSWEOAAABAVORARDSVDDIRVARVIEQDMVDSAGKITVRI 60

DB 1 MGHHHHHH----- 8

QY 61 KLEVSFKRPAQPRCKSPGSPETGAGAGTVATTPASSPVTLAETGTLTYPLFNLWG 120

DB 9 -----GSPFSGSPETGAGAGTVATTPASSPVTLAETGTLTYPLFNLWG 53

QY 121 PAFHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAKGLMNLALISA 180

DB 54 PAFHERYPNVITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAKGLMNLALISA 113

QY 181 QOVNYNLPGVSEHLKNGKLVIAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGS 240

DB 114 QOVNYNLPGVSEHLKNGKLVIAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDGS 173

QY 241 GDTFLFTQVLSKODPEGWCKSPGFGTTVDPPAVPGALGENGNGMVTGCAETPGCVAYIG 300
DB 174 GDTFLFTQVLSKODPEGWCKSPGFGTTVDPPAVPGALGENGNGMVTGCAETPGCVAYIG 233
QY 301 ISFLDQASQSGLEAQLGNSNGNFFLPDAQSIOAAAAAGFASKTPANQAISMIDGPAPDGY 360
DB 234 ISFLDQASQSGLEAQLGNSNGNFFLPDAQSIOAAAAAGFASKTPANQAISMIDGPAPDGY 293
QY 361 PIINYEYAIVNNRQKDAATAQTLQAFHWAITDGNKASFLDQVHFPOPLPVPVVKLSDALI 420
DB 294 PIINYEYAIVNNRQKDAATAQTLQAFHWAITDGNKASFLDQVHFPOPLPVPVVKLSDALI 353
QY 421 ATISSAEMKTDAAATLAQEAQGNPERISGDLKTQIDQVESTAGSLQGQWRAAGTAQAQAVV 480
DB 354 ATISSGG-----GSGGGSGGSGSVPTTAASPPSTA-----AAPPATPTVA 396
QY 481 RFQEAANKQKQELDEISTNIRQAGVOYVSRADBEQQOALSSQMGFTQSTQTVTDVQBEILNR 540
DB 397 PPPPAA-----ANTPNAQPG----- 411
QY 541 ANEVEAPMADPPPTDVPITPCELTAANKAAQQLVLSADN-----MREYLAAGAKERQ----- 591
DB 412 -----DPNAAAPPADPNAPPPVIAPNAPQPVRI--DNVGGFSFALPAGWVESDAHFD 454
QY 592 -----RLATSLRNAKAYGEVDEEAATALDND-- 618
DB 465 YGSALLSKTTGDPFPFPGQPPVANDTRIVLGRLDQKLYASAEA---TDSKAAARLGSMDG 521
QY 619 -----GEGTQVQASAGAVGDSABELTDTTPRVATAGEPNMDLKEAARKLETGD 667
DB 522 EYMPYVPGTRINQETVSLDANGVSSGSAAYEV-----KFSDFSPKNGQIWTGV 569
QY 668 QGASLAHFADGWMNTFNLTQGDVDRFRGDNWEGDAATACEASLDQORQWILHMAXLSAA 727
DB 570 IGSPAAANAPDA-----GPFQRW--FWVLGTANNPVDK----- 600
QY 728 MAKQAOYVLAQLHVWARREHPTVEDIVGLERLYAENPSARDQILPVYAEVQORSEKVLTEY 787
DB 601 -----GAAKALAE--SIRPLVAP----- 616
QY 788 NKAALPEVNNRKPPEPAIKIDPPPPQEQGLIFGLMPPSDGSGVTGTCMPAAPVPPPT 847
DB 617 -----PPAPAPA-PAEPAPAPAPAGEV-----AP-TPPT 643
QY 848 GSPGGGLPA 856
DB 644 PTPORTLPA 652

RESULT 7

US-08-818-112-153
Sequence 153, Application US/08818112
Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: PatentIn Release #1.0, Version #1.30
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/818,112
, FILING DATE: 13-MAR-1997
, CLASSIFICATION: 424
, ATTORNEY/AGENT INFORMATION:
, NAME: Makl, David J.
, REGISTRATION NUMBER: 31,392
, REFERENCE/DOCKET NUMBER: 210121.411C6
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (206) 622-4900
, TELEFAX: (206) 682-6031
, INFORMATION FOR SEQ ID NO: 153:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 374 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, US-08-818-112-153

```

Query Match	36.1%;	Score 1831;	DB 3;	Length 374;
Best Local Similarity	95.2%;	Pred. No. 1.3e-124;		
Matches 355;	Conservative 3;	Mismatches 1;	Indels 6;	Gaps 1;
QY	59	RIKLEVSFKMPAPQ-----RCGSKPPSGPSTGAGAGTVAITTPASSPVTLAETGSPFL 112		
Db	2	KIRLHTLAVTAAPLLAAAGCGSKPPSGPSTGAGAGTVAITTPASSPVTLAETGSPFL 61		
QY	113	YPLFNLGPAFHERYPNVITTAQGTSGGAGIAQAAAGTYNIGASDAYLSEGDMAAHKGLM 172		
Db	62	YPLFNLGPAFHERYPNVITTAQGTSGGAGIAQAAAGTYNIGASDAYLSEGDMAAHKGLM 121		
QY	173	NIALAISAQVNYNLPGVSEHLKLVLAAMTQGTIKTWDDPQIAALNPGVNLPGTAVV 232		
Db	122	NIALAISAQVNYNLPGVSEHLKLVLAAMTQGTIKTWDDPQIAALNPGVNLPGTAVV 181		
QY	233	PLHRSDGSGDTFLFTQYLSKQDPEGWKSFGFTTVDFAVPFALGNGNGGWTGCAET 292		
Db	182	PLHRSDGSGDTFLFTQYLSKQDPEGWKSFGFTTVDFAVPFALGNGNGGWTGCAET 241		
QY	293	PCGVAYIGISFLDQASORGICEAQLGNSSGNFLLPDAQSIQAAAAFPASKTPANQAISMI 352		
Db	242	PCGVAYIGISFLDQASORGICEAQLGNSSGNFLLPDAQSIQAAAAFPASKTPANQAISMI 301		
QY	353	DGPAPDGPVPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFILDQVHFQPLPPAV 412		
Db	302	DGPAPDGPVPIINYEYAIVNNRQKDAATAQTLQAFHLWAITDGNKASFILDQVHFQPLPPAV 361		
QY	413	VKLSDALIATISS 425		
Db	362	VKLSDALIATISS 374		

RESULT 8
 US-08-818-111-148
 ; Sequence 148, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle

```

; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-818-111-148

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 5; Gaps 1;

QY 59 RKLKLVSPKMPAPQ-----RCGKPPGSPPTGAGAGTVA--TTPASSPVLTAE--TGGSTLL 112
Db 2 KIRLHTLLAVLTAAFLPALLAAAGCGSKPPGSPPTGAGAGTVA--TTPASSPVLTAE--TGGSTLL 61
QY 113 YPLFNLWGPAPHERYPNVNVTITTAQCTSGAGIAQAAAGTVMIGASDAYI--SEGDMAAHKGLM 172
Db 62 YPLFNLWGPAPHERYPNVNVTITTAQCTSGAGIAQAAAGTVMIGASDAYI--SEGDMAAHKGLM 121
QY 173 NIALAISAAQVNNYLPVGSBELKINGVLAAAMYQGTITKWDPDFQIAALNPGVNLPGTAVV 232
Db 122 NIALAISAAQVNNYLPVGSBELKINGVLAAAMYQGTITKWDPDFQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDGSGDTFLFTQVLYSKQDPPEGWKSPGFGTTVDPPAVPGALGKNGNGMWGTGCAET 292
Db 182 PLHRSDGSGDTFLFTQVLYSKQDPPEGWKSPGFGTTVDPPAVPGALGKNGNGMWGTGCAET 241
QY 293 PCGVAYIGISFLDQASQRLGEAQLGNSSGNFLLPDAQSIQAAAAAGFASKTPANQAISMI 352
Db 242 PCGVAYIGISFLDQASQRLGEAQLGNSSGNFLLPDAQSIQAAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGPPIINYEYAVINNRQDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 412
Db 302 DGPAPDGPPIINYEYAVINNRQDAATAQTQLAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

```

```

; RESULT 9
; US-09-056-556-153
; Sequence 153, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue

```

TREATMENT

```

; CITY: ~ Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-056-556-153

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKRRPAQP-----RCGSKPPSGSPETGAGAGTATTTPASSPVTLLAETGSTALL 112
Db 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTATTTPASSPVTLLAETGSTALL 61
QY 113 YPLENLWGPAFHERYPNVTTTAQCTGSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 172
Db 62 YPLENLWGPAFHERYPNVTTTAQCTGSGAGIAQAAAGTNNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVNNLPGVSEHLKNGKVLAAVQGTIKTWDDPQIAALNPGVNLPGTAVV 232
Db 122 NIALAISAOQVNNLPGVSEHLKNGKVLAAVQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTLYQLSKQDPEGWGKSPGFGTTVDPPAVPGALGENGNGMVTGCAET 292
Db 182 PLHRSDSGDTFTLYQLSKQDPEGWGKSPGFGTTVDPPAVPGALGENGNGMVTGCAET 241
QY 293 PGCVAYIGISFLDQASQSGGLGSAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 352
Db 242 PGCVAYIGISFLDQASQSGGLGSAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAI SMI 301
QY 353 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 412
Db 302 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

```

```

RESULT 10
US-09-056-556-155
; Sequence 155, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
TREATM

```

```

RESULT 11
US-09-072-596-148
; Sequence 148, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.

```

APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/072,596
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-596-148

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
QY 59 RIKLEVSFKVRPAQP-----RCGSKPPSPGSGAGAGTAVTTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGPETGAGTAVTTPASSPVTLAETGSTALL 61
QY 113 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVYNLPGVSEHLKNGVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPGVSEHLKNGVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTLFTQYLSKQDPEGWKGSPFGTGTVDFFAVPGALGNGGNGVTCGAET 292
DB 182 PLHRSDSGDTFTLFTQYLSKQDPEGWKGSPFGTGTVDFFAVPGALGNGGNGVTCGAET 241
QY 293 PGCVAYIGISFLDQASORGLEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASORGLEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIIVNRRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIIVNRRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

RESULT 12

US-09-072-596-150

; Sequence 150, Application US/09072596

; Patent No. 6458366

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Twardzik, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/072,596
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 150:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-150

Query Match 36.1%; Score 1831; DB 4; Length 374;
Best Local Similarity 95.2%; Pred. No. 1.3e-124;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
QY 59 RIKLEVSFKVRPAQP-----RCGSKPPSPGSGAGAGTAVTTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGPETGAGTAVTTPASSPVTLAETGSTALL 61
QY 113 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLENLGPAFHRYPNVTTTAQGTSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVYNLPGVSEHLKNGVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPGVSEHLKNGVLAAMYQGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDSGDTFTLFTQYLSKQDPEGWKGSPFGTGTVDFFAVPGALGNGGNGVTCGAET 292
DB 182 PLHRSDSGDTFTLFTQYLSKQDPEGWKGSPFGTGTVDFFAVPGALGNGGNGVTCGAET 241
QY 293 PGCVAYIGISFLDQASORGLEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAYIGISFLDQASORGLEAQLGNSGNFLLPDAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAIIVNRRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAIIVNRRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

Db 362 VKLSDALIATISS 374

RESULT 13

US-09-118-426-5

Sequence 5, Application US/09118426C

Patent No. 6517839

GENERAL INFORMATION:

APPLICANT: Modlin, Robert L.

APPLICANT: Libraty, Daniel H.

TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

TITLE OF INVENTION: T-CELL RESPONSE

FILE REFERENCE: 30435.4US01

CURRENT APPLICATION NUMBER: US/09/118,426C

EARLIER FILING DATE: 1998-07-17

EARLIER FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 373

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

PUBLICATION INFORMATION:

JOURNAL: Infect. Immun.

VOLUME: 57

ISSUE: 8

PAGES: 2481-

DATE: 1989

US-09-118-426-5

Query Match 36.1%; Score 1829; DB 4; Length 373;

Best Local Similarity 100.0%; Pred. No. 1.8e-124;

Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 GSKPPSPGSPETGAGAGTVAATTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 134

Db 23 GSKPPSPGSPETGAGAGTVAATTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 82

QY 135 QGTGSGAGTAQAAGTGNIGASDAYLSEGDMAAHKGLMNLALISAQOVNPNLPGVSEHL 194

Db 83 QGTGSGAGTAQAAGTGNIGASDAYLSEGDMAAHKGLMNLALISAQOVNPNLPGVSEHL 142

QY 195 KLGKVLAAMYOGTITKWDPPQIAALNPGVNLPGTAVVPLHRSKQD 254

Db 143 KLGKVLAAMYOGTITKWDPPQIAALNPGVNLPGTAVVPLHRSKQD 202

QY 255 PEGWKSFGTITVDPFPAVPGALGENGNGMVTGCAETPGCVAYIGISFLDQASORGLGE 314

Db 203 PEGWKSFGTITVDPFPAVPGALGENGNGMVTGCAETPGCVAYIGISFLDQASORGLGE 262

QY 315 AOLGNSSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAIDGPGYPIINYEYAIYNNRQ 374

Db 263 AOLGNSSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAIDGPGYPIINYEYAIYNNRQ 322

QY 375 KDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALIATISS 425

Db 323 KDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALIATISS 373

RESULT 14

US-09-118-426-6

Sequence 6, Application US/09118426C

Patent No. 6517839

GENERAL INFORMATION:

APPLICANT: Modlin, Robert L.

APPLICANT: Libraty, Daniel H.

TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

TITLE OF INVENTION: T-CELL RESPONSE

FILE REFERENCE: 30435.4US01

CURRENT APPLICATION NUMBER: US/09/118,426C

EARLIER FILING DATE: 1998-07-17

EARLIER FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 351

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

PUBLICATION INFORMATION:

JOURNAL: Infect. Immun.

VOLUME: 57

ISSUE: 8

PAGES: 2481-

DATE: 1989

US-09-118-426-6

Query Match 35.9%; Score 1820; DB 4; Length 351;

Best Local Similarity 100.0%; Pred. No. 7.4e-124;

Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 GSKPPSPGSPETGAGAGTVAATTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 135

Db 2 GSKPPSPGSPETGAGAGTVAATTPASSPVTLAETGTLTLLYPLNMGPAFHRYPNVTITA 61

QY 136 GTGSGAGTAQAAGTGNIGASDAYLSEGDMAAHKGLMNLALISAQOVNPNLPGVSEHLK 195

Db 62 GTGSGAGTAQAAGTGNIGASDAYLSEGDMAAHKGLMNLALISAQOVNPNLPGVSEHLK 121

QY 196 LNKVLAAMYOGTITKWDPPQIAALNPGVNLPGTAVVPLHRSKQD 255

Db 122 LNKVLAAMYOGTITKWDPPQIAALNPGVNLPGTAVVPLHRSKQD 181

QY 256 EGWKSFGTITVDPFPAVPGALGENGNGMVTGCAETPGCVAYIGISFLDQASORGLGE 315

Db 182 EGWKSFGTITVDPFPAVPGALGENGNGMVTGCAETPGCVAYIGISFLDQASORGLGE 241

QY 316 QLGNSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAIDGPGYPIINYEYAIYNNRQ 375

Db 242 QLGNSGNFLPDQAQSIQAAAAGFASKTPANQAISMIDGPAIDGPGYPIINYEYAIYNNRQ 301

QY 376 DAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALIATISS 425

Db 302 DAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAVVKSALIATISS 351

RESULT 15

US-08-818-112-89

Sequence 89, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedwick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-89

Query Match      15.6%; Score 791; DB 3; Length 166;
Best Local Similarity 95.2%; Pred. No. 7.4e-50;
Matches 157; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 525 TQQTIVTDQOEILNRANEVEAPMDPPTDVPITPCELTAAKNAQQLVLSADNMREYLA 584
Db 2 TQQTIVTDQOEILNRANEVEAPMDPPTDVPITPCELTXXKNAQQLVLSADNMREYLA 61

Qy 585 AGAKERORLATSRLNAAKAYGEVDEEAATALDNDGEGTVQAESAGAVGGSSAELTDTPR 644
Db 62 AGAKERORLATSRLNAAKAYGEVDEEAATALDNDGEGTVQAESAGAVGGSSAELTDTPR 121

Qy 645 VATAGEPNFMDLKEAAKLETGQOGASLAHFADGWNFTNLTQGD 689
Db 122 VATAGEPNFMDLKEAAKLETGQOGASLAHXGDGWNFTNLTQGD 166

Search completed: November 21, 2003, 16:11:43
Job time : 22.7299 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 38.3392 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHHHHHHVIDIIGTSPTW.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2700	53.2	802	9	US-09-287-849-10
2	2700	53.2	802	12	US-10-084-843-214
3	2700	53.2	802	12	US-10-084-843-351
4	2700	53.2	802	12	US-10-193-002-209
5	2700	53.2	802	12	US-10-193-002-346
6	2700	53.2	802	12	US-10-359-460-10
7	2375	46.8	460	12	US-10-084-843-184
8	2375	46.8	460	12	US-10-193-002-179
9	1839	36.3	652	12	US-10-084-843-355
10	1839	36.3	652	12	US-10-193-002-350
11	1831	36.1	374	9	US-09-287-849-6
12	1831	36.1	374	9	US-09-287-849-40
13	1831	36.1	374	12	US-10-084-843-153
14	1831	36.1	374	12	US-10-084-843-155
15	1831	36.1	374	12	US-10-193-002-148

16	1831	36.1	374	12	US-10-193-002-150
17	1831	36.1	374	12	US-10-359-460-6
18	1831	36.1	374	12	US-10-359-460-40
19	1831	36.1	374	12	US-10-098-732A-39
20	791	15.6	166	12	US-10-084-843-89
21	791	15.6	166	12	US-10-193-002-90
22	485	9.6	100	12	US-10-084-843-115
23	485	9.6	100	12	US-10-193-002-110
24	485	9.6	100	16	US-10-080-170-639
25	481	9.5	100	15	US-10-140-045-5
26	479.5	9.5	358	9	US-09-287-849-8
27	479.5	9.5	358	12	US-10-359-460-8
28	466	9.2	95	12	US-10-084-843-88
29	466	9.2	95	12	US-10-193-002-89
30	466	9.2	95	12	US-10-098-732A-35
31	413	8.1	346	9	US-09-741-669-380
32	392	7.7	80	12	US-10-084-843-117
33	392	7.7	80	12	US-10-193-002-112
34	329.5	6.5	364	12	US-10-084-843-73
35	329.5	6.5	364	12	US-10-193-002-74
36	324	6.4	66	12	US-10-084-843-77
37	324	6.4	66	12	US-10-193-002-78
38	324	6.4	66	12	US-10-098-732A-37
39	313.5	6.2	368	15	US-10-156-761-11606
40	276	5.4	375	10	US-09-738-636-6333
41	236	4.7	49	15	US-10-140-045-6
42	217.5	4.3	1463	10	US-09-971-536-69
43	205	4.0	42	15	US-10-140-045-7
44	202.5	4.0	2468	12	US-10-246-330-4
45	192	3.8	856	9	US-09-287-849-12

ALIGNMENTS

RESULT 1

US-09-287-849-10
Sequence 10, Application US/09287849
Patent No. US20020009459A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-003020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 802
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10

Query Match 53.2% Score 2700; DB 9; Length 802;
Best Local Similarity 66.6%; Pred. No. 6.1e-161;

```
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHVHHVHVDIIGTSTSEWQAAAVQARSDVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHVHHVHVDIIGTSTSEWQAAAVQARSDVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 120
Db 61 KLEVSFMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 119
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLGNIAALISA 180
Db 120 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLGNIAALISA 179
QY 181 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
QY 241 GDTFLFTQYLSKODPEGWKGSPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKGSPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 299
QY 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAAGFASKTPANQAISMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAAGFASKTPANQAISMIDGPADGY 359
QY 361 PIINYEYAI VNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQPLPPAVVKS DALI 420
Db 360 PIINYEYAI VNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQPLPPAVVKS DALI 419
```

RESULT 2

```
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;             Skeiky, Yasir A.W.
;             Dillon, Davin C.
;             Campos-Neto, Antonio
;             Houghton, Raymond
```

```
;
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
;
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
;
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
;
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ. ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 214:
;
; US-10-084-843-214
```

```
Query Match 53.28; Score 2700; DB 12; Length 802;
Best Local Similarity 66.6%; Pred. No. 6.1e-161;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;
QY 1 MGHVHHVHVDIIGTSTSEWQAAAVQARSDVDDIRVARVIEQDMVDSAGKITRYI 60
Db 1 MGHVHHVHVDIIGTSTSEWQAAAVQARSDVDDIRVARVIEQDMVDSAGKITRYI 60
QY 61 KLEVSFMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 120
Db 61 KLEVSFMRPAQRCGSKPPSGSPETGAGAGTATTTPASSPVTLAETGTLTLLPLFNLWG 119
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLGNIAALISA 180
Db 120 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTNTGASDAYLSEGDMAHKLGNIAALISA 179
QY 181 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
Db 180 QQVNNYLPVSEHLKNGKVLAAAMYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
QY 241 GDTFLFTQYLSKODPEGWKGSPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 300
Db 240 GDTFLFTQYLSKODPEGWKGSPGFTTVPFPAVPGALGENGGMVTCGAETPGCVAYIG 299
QY 301 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAAGFASKTPANQAISMIDGPADGY 360
Db 300 ISFLDQASQRLGEAQLGNSSGNFLPDAQSIQAAAAGFASKTPANQAISMIDGPADGY 359
QY 361 PIINYEYAI VNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQPLPPAVVKS DALI 420
Db 360 PIINYEYAI VNNRQKDAATAQTLQAEFLHWAITDGNKASFLOVHFQPLPPAVVKS DALI 419
```



```

QY 421 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 480
DB 420 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 479
QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQVTVDQOEILNR 540
DB 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQVTVDQOEILNR 537
QY 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQLATSLRNA 600
DB 538 PPAPATPVAPPPPPAAANTFNAQPGDPAAPPPADPNAPPVPIAPNAPQVPR----- 589
QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SALETDTPRVATAGEPNF- 653
DB 590 -----IDNPVGGFSFALPAGWVESDAAHFYDGSALLS-----KITGDPFPP 630
QY 654 -----MDLKEAARKLETGDOGASLAHFAUGWNTFNLTLOGDVKRRFGFD 697
DB 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFEF--YM 675
QY 698 NWEGDAATACEASLDQQRQWILHMAKLSAAMAKQAQYVAQLHVMARREHPTIEDIVGLER 757
DB 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKENGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPYAYEQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQRFVFW-----LGTANNPVDKGAALAESIRPLVAPPPA-----P 770
QY 810 PPPQEQGLIFGLMPPSDSGVTPGTGMPAPMVPTGSPCGGLPA 856
DB 771 APAPAEPA-----PAPAPAGEVAP-----TPPTPTPQRTLPA 802

RESULT 3
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;             Skeiky, Yasir A.W.
;             Dillon, David C.
;             Campos-Neto, Antonio
;             Houghton, Raymond
;             Vedvick, Thomas S.
;             Twardzik, Daniel R.
;             Lodes, Michael J.
;             Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/084,843
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392

```

```

; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 351:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match      53.2%; Score 2700; DB 12; Length 802;
Best Local Similarity 66.6%; Pred. No. 6.1e-161;
Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHHHHHHVIDIGTSPTSEWQAAAPAVQARSDSVDDIRVARVIEQDMAVDSAGKITRYI 60
DB 1 MGHHHHHHVIDIGTSPTSEWQAAAEAVQARSDSVDDIRVARVIEQDMAVDSAGKITRYI 60
QY 61 KLEVSFKMPPAOPRCGSKPPSGPETGAGCTVATTPASSPVTLAETGSLTLYPLFNMG 120
DB 61 KLEVSFKMPPAOPRR-GSKPPSGPETGAGCTVATTPASSPVTLAETGSLTLYPLFNMG 119
QY 121 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMKNIALAISA 180
DB 120 PAFHERYPNVTTTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLMKNIALAISA 179
QY 181 QQVNNYLPGVSEHLKLVLAAMVQGTITKTWDDPQIAALNFGVNLPGTAVVPLHRSDGS 240
DB 180 QQVNNYLPGVSEHLKLVLAAMVQGTITKTWDDPQIAALNFGVNLPGTAVVPLHRSDGS 239
QY 241 GDTFLTQVLSKQDPGCKSPGFGTTVDPPAVPGALGENGMVGTGCAETPGCVAYIG 300
DB 240 GDTFLTQVLSKQDPGCKSPGFGTTVDPPAVPGALGENGMVGTGCAETPGCVAYIG 299
QY 301 ISFLDAQSORGLGEAQIIGNSSGNFLLPDAQSIQAAAAGFASKTPANQA-SMIDGPADGY 360
DB 300 ISFLDAQSORGLGEAQIIGNSSGNFLLPDAQSIQAAAAGFASKTPANQAISMIDGPADGY 359
QY 361 PIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDOVHFOPLPVPVVKLSDALI 420
DB 360 PIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDOVHFOPLPVPVVKLSDALI 419
QY 421 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 480
DB 420 ATISSAEMKTDATLAQEAAGNFERISGDLTKTQIDQVESTAGSLOGQREGAGTAQAQAVV 479
QY 481 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQVTVDQOEILNR 540
DB 480 RFOEAANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGFTQSQVTVDQOEILNR 537
QY 541 ANEVEAPMADPPTDVPITPCELTAAKNAQQLVLSADNMREYLAAGAKERQLATSLRNA 600
DB 538 PPAPATPVAPPPPPAAANTFNAQPGDPAAPPPADPNAPPVPIAPNAPQVPR----- 589
QY 601 AKAYGEVDEEAATALDNDGEGTVQAESAGAVGDS-----SALETDTPRVATAGEPNF- 653
DB 590 -----IDNPVGGFSFALPAGWVESDAAHFYDGSALLS-----KITGDPFPP 630
QY 654 -----MDLKEAARKLETGDOGASLAHFAUGWNTFNLTLOGDVKRRFGFD 697
DB 631 GQPPPVANDTRIVLGRDQKLYASAEATDSKAAA-----RLGSDMGFEF--YM 675
QY 698 NWEGDAATACEASLDQQRQWILHMAKLSAAMAKQAQYVAQLHVMARREHPTIEDIVGLER 757
DB 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFSKENGQIWTGVIGSPA 724
QY 758 LYAENPSARDQILPYAYEQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
DB 725 ANAPDAGPPQRFVFW-----LGTANNPVDKGAALAESIRPLVAPPPA-----P 770

```

QY 810 PPPQOGLIPGLMPPSDGSGVTPTGMPAABWVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPORTLPA 802

RESULT 4

US-10-193-002-209
 ; Sequence 209, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/193,002
 ; FILING DATE: 10-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 209:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 802 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 209:
 ; US-10-193-002-209

Query Match 53.2%; Score 2700; DB 12; Length 802;
 Best Local Similarity 66.8%; Pred. No. 6.1e-161;
 Matches 591; Conservative 27; Mismatches 153; Indels 116; Gaps 15;

QY 1 MGHVHHHHVIDIIGTSTWEQAAAEAVQARDSDIRVARVIEDQMAVDSAGKITVRI 60
 Db 1 MGHVHHHHVIDIIGTSTWEQAAAEAVQARDSDIRVARVIEDQMAVDSAGKITVRI 60
 QY 61 KLVSVFMRPAQRCCKSPSPGSPETGAGAGVATTPASSPTVLAETGSLTLLYPLFNWLG 120
 Db 61 KLVSVFMRPAQRCCKSPSPGSPETGAGAGVATTPASSPTVLAETGSLTLLYPLFNWLG 119
 QY 121 PAFHRYPNVTTAQTGSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLMNIALAISA 180
 Db 120 PAFHRYPNVTTAQTGSGAGIAQAAGTAVNIGASDAYLSEGDMAAHKGLMNIALAISA 179

QY 181 QCVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 240
 Db 180 QCVNYNLPGVSEHLKNGKVLAAVYQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRS DGS 239
 QY 241 GDTFLFTOYLSKQDPEGKSPGFGTTVDFFPAVPGALGENGNGMWTCGAETPGCVAYIG 300
 Db 240 GDTFLFTOYLSKQDPEGKSPGFGTTVDFFPAVPGALGENGNGMWTCGAETPGCVAYIG 299
 QY 301 ISFLDQASQRLGEAOLGNSSGNFLPDQOSIQAAAGAFASKTPANQALSMIDGPADGY 360
 Db 300 ISFLDQASQRLGEAOLGNSSGNFLPDQOSIQAAAGAFASKTPANQALSMIDGPADGY 359
 QY 361 PIINYEYAI VNNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 420
 Db 360 PIINYEYAI VNNRQKDAATQTLQAFHLHWAITDGNKASFLDQVHFQPLPFAVVKLS DALI 419
 QY 421 ATISSAEMKTDAAATLAQEAENFERISGDLKTQIDQVESTAGSLQGWGRGAAGTAAQAAV 480
 Db 420 ATISSAEMKTDAAATLAQEAENFERISGDLKTQIDQVESTAGSLQGWGRGAAGTAAQAAV 479
 QY 481 RFQEAANKQKQELDEISTNIRQAGVQYSRADDEEQOQALSSQMGFTQSTVTVDQOEILNR 540
 Db 480 RFQEAANKQKQELDEISTNIRQAGVQYSRADDEEQOQALSSQMGFTQSTVTVDQOEILNR 537
 QY 541 ANEVEAPMADPTDVTPTCELTAAKVAQOLVLSADNMREYLAAGAKERQRLATSLRNA 600
 Db 538 PPAPATPVAPPPPPAAAANTPNAQPCDPNAPPPADPNAPPPPPPIAPNAPQVPR----- 589
 QY 601 AKAYGEVDEEAATALDNDGEGTVCAESAGAVGDS-----SALETDTPRVATAGBNP- 653
 Db 590 -----IDNPVGGSFALPAGWVESDAAHFDYGSALLS-----KTTGDPPPP 630
 QY 654 -----MDLXEAARKLETGQOGASLAHFAFDGWNFTNLTLOGDVYKFRGPD 697
 Db 631 GQPPPVANDTRIVGLRDLQKLYASAEATDSKAAA-----RLGSDMGEF--YM 675
 QY 698 NWECDATAACEASLDQORQWILHMAKLSAAMAKQAVQALHVWARREHTYEDIVGLER 757
 Db 676 PYPGTRINQETVSLD-----ANGVSGSASYEVKFSDFPSKPNQINTGTGIGSPA 724
 QY 758 LYAENPSARDQILPVVYAEYQORSEKVLTEYNN-----KAALEPVNP-PKPPPAIKIDP 809
 Db 725 ANAFDAGPPQRFVFW-----LGTANNPVDKGAALAEISIRPLVAPPPA--P 770
 QY 810 PPPQOGLIPGLMPPSDGSGVTPTGMPAABWVPTGSPGGGLPA 856
 Db 771 APAPAEPA-----PAPAPAGEVAP-----TPTTPTPORTLPA 802

RESULT 5

US-10-193-002-346
 ; Sequence 346, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 179:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 179:
US-10-193-002-179
Query Match 46.8%; Score 2375; DB 12; Length 460;
Best Local Similarity 100.0%; Pred. No. 6.9e-141;
Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 525 TQSQTVTVDQBIILNRAVEAPMADPTDPTIPCELTAAKNAQAQVLVSADNNREYLA 584
DB 2 TQSQTVTVDQBIILNRAVEAPMADPTDPTIPCELTAAKNAQAQVLVSADNNREYLA 61
QY 585 AGAKERQRLATSLRNAKAYGVDEEATAALDNDGEGTVQAESAGAVGDSASALTDPTR 644
DB 62 AGAKERQRLATSLRNAKAYGVDEEATAALDNDGEGTVQAESAGAVGDSASALTDPTR 121
QY 645 VATAGEPNFMDLKEARKLETQDQASLAHFDAGWNTNLTQGDVKEFRGFDNWECDAA 704
DB 122 VATAGEPNFMDLKEARKLETQDQASLAHFDAGWNTNLTQGDVKEFRGFDNWECDAA 181
QY 705 TACEASLDQOROWILHMAKLSAAMAKQAYVAQLHWARREHTPTVEDIVGERLYAENPS 764
DB 182 TACEASLDQOROWILHMAKLSAAMAKQAYVAQLHWARREHTPTVEDIVGERLYAENPS 241
QY 765 ARDQILPVVAYEQORSEKVLTEYNNKAALPEVNPPEKPPPAIKIDPPPPPPQEQGLIPGLM 824
DB 242 ARDQILPVVAYEQORSEKVLTEYNNKAALPEVNPPEKPPPAIKIDPPPPPPQEQGLIPGLM 301
QY 825 PRSDGSGVTPTGMPAAPVWPTGSPGGGLPADTAQAQLTSAGREAAALSGDVAVKAASLG 884

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355
Query Match 36.3%; Score 1839; DB 12; Length 652;
Best Local Similarity 48.5%; Pred. No. 4.4e-107;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHHVIIDTGTGTSWEQAABAVQARDSDIRVARVIEQDMAVDSAGKITRYI 60
DB 1 MGHHHHHH-----8
QY 61 KLEVSFKMRPAQPRCGSKPPSGFETGAGAGTVAATTPASSPVTLAETGTLPLPLNLG 120

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355
Query Match 36.3%; Score 1839; DB 12; Length 652;
Best Local Similarity 48.5%; Pred. No. 4.4e-107;
Matches 441; Conservative 26; Mismatches 132; Indels 310; Gaps 19;
QY 1 MGHHHHHHVIIDTGTGTSWEQAABAVQARDSDIRVARVIEQDMAVDSAGKITRYI 60
DB 1 MGHHHHHH-----8
QY 61 KLEVSFKMRPAQPRCGSKPPSGFETGAGAGTVAATTPASSPVTLAETGTLPLPLNLG 120

[illegible]

RESULT 11

```

US-09-287-849-6
; Sequence 6. Application US/09287849
; Patent No. US20020009459A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
;
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
;
; TITLE OF INVENTION: and their Uses
;
; FILE REFERENCE: 014058-009020US
;
; CURRENT APPLICATION NUMBER: US/09/287,849
;
; CURRENT FILING DATE: 1999-04-07
;
; PRIOR APPLICATION NUMBER: US 08/818,112
;
; PRIOR FILING DATE: 1997-03-13
;
; PRIOR APPLICATION NUMBER: US 08/942,578
;
; PRIOR FILING DATE: 1997-10-01
;
; PRIOR APPLICATION NUMBER: US 09/025,197
;
; PRIOR FILING DATE: 1998-02-18
;
; PRIOR APPLICATION NUMBER: US 09/056,556
;
; PRIOR FILING DATE: 1998-04-07
;
; PRIOR APPLICATION NUMBER: US 09/223,040
;
; PRIOR FILING DATE: 1998-12-30
;
; NUMBER OF SEQ ID NOS: 46
;
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 5
;
; LENGTH: 374
;
; TYPE: PRT
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-6

```

Query Match

374;

[illegible]

RESULT 12

```

US-09-287-849-40
; Sequence 40, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Protiens of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kd antigen
US-09-287-849-40

```

Query Match 36.1%; Score 1831; DB 9; Length 374;

Db	2	KIRLHLLAVLTAAPLFLIAAAGCGSKPPSGPETGAGAGTAVTPASSPVTLAETGSTLL	61
Qy	113	YFLFNLWGPAPHERYPNVTTITAQGTGSGAGIAQAAAGTVNIGASDAYLISEGDMAAHKGLM	172
Db	62	YFLFNLWGPAPHERYPNVTTITAQGTGSGAGIAQAAAGTVNIGASDAYLISEGDMAAHKGLM	121
Qy	173	NIALAISAQVNVNYPGVSEHLKUNGKVLAMYGOTTKTWDDPQIALALNPGVNLPGTAV	232
Db	122	NIALAISAQVNVNYPGVSEHLKUNGKVLAMYGOTTKTWDDPQIALALNPGVNLPGTAV	181
Qy	233	PLHRSDGSGDTFLFTQYLSKODPEGWGKSPQFGFTTVDFPAPVPCALGENGGWMTGCAS	292
Db	182	PLHRSDGSGDTFLFTQYLSKODPEGWGKSPQFGFTTVDFPAPVPCALGENGGWMTGCAS	241
Qy	293	PGCVAYIGISFLDQASQRGLGEALGNSSGNFLLPDPAQSTQAAAAAGFASKTPANQAISMI	352
Db	242	PGCVAYIGISFLDQASQRGLGEALGNSSGNFLLPDPAQSTQAAAAAGFASKTPANQAISMI	301
Qy	353	DGAPDGPPIINYEYAIYVNNRQKDAATAQTQLAELHWAITDGNKASFLDQVHFQPLPPAV	412
Db	302	DGAPDGPPIINYEYAIYVNNRQKDAATAQTQLAELHWAITDGNKASFLDQVHFQPLPPAV	361
Qy	413	VKLSDALIATISS	425
Db	362	VKLSDALIATISS	374

RESULT 13

US-10-084-843-153
Sequence 153, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:

[illegible]

RESUIT 14

RESULTS 14
US-10-084-843-155
Sequence 155, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Steiky, Yasir A.W.
Dillon, Davin C.
Campos-Neco, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSED: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/10/084,843
APPLICATION DATA:
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 155:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 155:
US-10-084-843-155

Query Match 36.1%; Score 1831; DB 12; Length 374;
Best Local Similarity 95.2%; Pred. No. 6.6e-107;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKMPAPQ-----RCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 61
QY 113 YPLFNLMGPAPHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLMGPAPHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVYNLPVSEHLKNGKVLAAAMYOGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPVSEHLKNGKVLAAAMYOGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDGSGDTFLTQYLSKQDPEGWKGSPGFGTTVDPPAVPGALGENGNGMVTGCAET 292
DB 182 PLHRSDGSGDTFLTQYLSKQDPEGWKGSPGFGTTVDPPAVPGALGENGNGMVTGCAET 241
QY 293 PGCVAIIGISFLDQASQRLGEAQLGNSGNGFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAIIGISFLDQASQRLGEAQLGNSGNGFLLPDAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

RESULT 15
US-10-193-002-148
Sequence 148, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/193,002
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-193-002-148

Query Match 36.1%; Score 1831; DB 12; Length 374;
Best Local Similarity 95.2%; Pred. No. 6.6e-107;
Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKMPAPQ-----RCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 112
DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSGSPETGAGAGTVAATTPASSPVTLAETGSTALL 61
QY 113 YPLFNLMGPAPHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 172
DB 62 YPLFNLMGPAPHERYPNVTITTAQGTGSGAGIAQAAAGTVNIGASDAYLSEGDMAAHKGLM 121
QY 173 NIALAISAOQVYNLPVSEHLKNGKVLAAAMYOGTIKTWDDPQIAALNPGVNLPGTAVV 232
DB 122 NIALAISAOQVYNLPVSEHLKNGKVLAAAMYOGTIKTWDDPQIAALNPGVNLPGTAVV 181
QY 233 PLHRSDGSGDTFLTQYLSKQDPEGWKGSPGFGTTVDPPAVPGALGENGNGMVTGCAET 292
DB 182 PLHRSDGSGDTFLTQYLSKQDPEGWKGSPGFGTTVDPPAVPGALGENGNGMVTGCAET 241
QY 293 PGCVAIIGISFLDQASQRLGEAQLGNSGNGFLLPDAQSIQAAAAGFASKTPANQAISMI 352
DB 242 PGCVAIIGISFLDQASQRLGEAQLGNSGNGFLLPDAQSIQAAAAGFASKTPANQAISMI 301
QY 353 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 412
DB 302 DGPAPDGYPIINYEYAVNNRQKDAATAQTLOAFLHWAITDGNKASFLDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
DB 362 VKLSDALIATISS 374

Search completed: November 21, 2003, 16:38:18
Job time : 42.3392 secs

This page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 20.507 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-54
Perfect score: 5072
Sequence: 1 MGHHHHHVIDIIGTSPTSW.....RAWTEAVIGNRRQDSKESK 983

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76.*

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2375	46.8	460	2 G70803	hypothetical prote
2	1831	36.1	374	1 F70584	phosphate specific
3	485	9.6	100	2 H70802	hypothetical prote
4	482.5	9.5	392	2 AG2377	phosphate-binding
5	463	9.1	364	2 G82595	ABC transporter ph
6	460.5	9.1	373	2 H72756	probable phosphate
7	456	9.0	347	2 AD1920	ABC phosphate tran
8	445	8.8	403	2 H75611	phosphate ABC tran
9	444	8.8	383	1 S74423	phosphate-binding
10	440	8.7	346	2 AI0499	probable phosphate
11	419	8.3	346	2 H91211	phosphate-binding
12	419	8.3	346	2 H65057	phosphate-binding
13	413	8.1	346	1 BYSCPR	phosphate-repressi
14	411	8.0	346	2 AB0956	periplasmic phosph
15	406	8.0	383	2 D75081	phosphate abc tran
16	404	8.0	389	1 F64426	phosphate-binding
17	400	7.9	337	1 C70473	phosphate-binding
18	380.5	7.5	405	2 C90194	hypothetical prote
19	378	7.5	429	2 C87183	phosphate-binding
20	335.5	6.6	369	2 B81171	Psst component of
21	330.5	6.5	370	2 D70584	phosphate uptake s
22	325	6.4	71	2 F70833	probable protein t
23	316.5	6.2	333	1 S74876	phosphate-binding
24	314	6.2	370	2 H70583	phosphate-binding
25	304	6.0	258	1 I64120	phosphate-binding
26	226.5	4.5	326	1 S39852	phosphate-binding
27	220	4.3	298	2 D97110	periplasmic phosph
28	203	4.0	1804	2 H96597	hypothetical prote
29	202.5	4.0	298	2 A86840	hypothetical prote

ALIGNMENTS

RESULT 1

G70803 hypothetical protein Rv3881c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: G70803

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence: a reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: G70803

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-460 <COL>

A/Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL1973.1; PID:el264132

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: Rv3881c

Query Match 46.8%; Score 2375; DB 2; Length 460;

Best Local Similarity 100.0%; Pred. No. 2.9e-106;

Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 525 TQSGTIVDQOEILNRANEVEAPMADPPTDVPITPCELTAAKNAQAQVLVSADNMREYLA 584

Db 2 TQSGTIVDQOEILNRANEVEAPMADPPTDVPITPCELTAAKNAQAQVLVSADNMREYLA 61

Qy 585 AGAKERORLATSRLNAAKAYGEVDEBAATLNDGEGTVOAESAGAVGGSSAELDTTPR 644

Db 62 AGAKERORLATSRLNAAKAYGEVDEBAATLNDGEGTVOAESAGAVGGSSAELDTTPR 121

Qy 645 VATAGEFNFDLKEARKLETGQGGASLAHFAAGWNTFNLTLOQDVKRPFQFNWEGDAA 704

Db 122 VATAGEFNFDLKEARKLETGQGGASLAHFAAGWNTFNLTLOQDVKRPFQFNWEGDAA 181

Qy 705 TACEASLDQQRQWILHMAKLSAAVAQAQYVAQLHVWARREHPTIEDIVGLEKLYAENPS 764

Db 182 TACEASLDQQRQWILHMAKLSAAVAQAQYVAQLHVWARREHPTIEDIVGLEKLYAENPS 241

Qy 765 ARDQILPVYAEYQORSEKVLTEYNNKALEPVNPKPPPAIKIDPPPPQOGLIPGFLM 824

Db 242 ARDQILPVYAEYQORSEKVLTEYNNKALEPVNPKPPPAIKIDPPPPQOGLIPGFLM 301

Qy 825 PPSDGSVTPGTGMPAAPMPVPTGSPGGGLPADTAQAQLTSAGREAAALSGDVAKKAASLG 894

Db 302 PPSDGSVTPGTGMPAAPMPVPTGSPGGGLPADTAQAQLTSAGREAAALSGDVAKKAASLG 361

Qy 885 GGGGGVPSAPLGSATGGAESVTPAGAGDTAGLQGGAGGAALGGGNGMPGAHQGQ 944

Db 362 GGGGGVPSAPLGAISGAESVRPAGAGDIAGLQGRAGGAALGGGCMGMPMAHQOQ 421

QY 945 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 983

Db 422 GGAKSQSGQDEALYTEDRAWTEAVIGNRRQDSKESK 460

RESULT 2

F70584

Phosphate specific transporter S precursor - Mycobacterium tuberculosis (strain H37Rv)

N;Alternate names: antigen b

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C;Accession: F70584; JCS103; A42930; A49721; A45820

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70584

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-374 <COL>

A;Cross-references: GB:295209; GB:AL123456; NID:g3261750; PIDN:CAB08484.1; PID:g2078049

A;Experimental source: strain H37Rv

R;Braibant, M.; Lefevre, P.; de Wit, L.; Pairs, P.; Ooms, J.; Huygen, K.; Andersen, A.B.; Gene 176, 171-176, 1996

A;Title: A Mycobacterium tuberculosis gene cluster encoding proteins of a phosphate transporter

A;Reference number: JCS100; MUID:97075926; PMID:8918249

A;Accession: JCS103

A;Molecule type: DNA

A;Residues: 1-374 <AND>

A;Cross-references: GB:M30046; NID:g149987; PIDN:AAA25374.1; PID:g149988

A;Note: neither the complete nucleic acid sequence nor the complete translation are shown

R;Andersen, A.B.; Hansen, E.B.; Infect. Immun. 57, 2481-2488, 1989

A;Title: Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular weight protein

A;Reference number: A42930; MUID:89307568; PMID:2545626

A;Accession: A42930

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-374 <AND1>

A;Cross-references: GB:M30046; NID:g149987; PIDN:AAA25374.1; PID:g149988

R;Chang, Z.; Choudhary, A.; Lathigra, R.; Quijcho, F.A.; J. Biol. Chem. 269, 1956-1958, 1994

A;Title: The immunodominant 38-kDa lipoprotein antigen of Mycobacterium tuberculosis is encoded by a single gene

A;Reference number: A49721; MUID:94124544; PMID:8294447

A;Accession: A49721

A;Molecule type: protein

A;Residues: 25-34 <CHA>

R;Andersen, A.B.; Ljungqvist, L.; Olsen, M.; J. Gen. Microbiol. 136, 477-480, 1990

A;Title: Evidence that protein antigen b of Mycobacterium tuberculosis is involved in phosphate transport

A;Reference number: A45820; MUID:90362031; PMID:2118164

A;Accession: A45820

A;Molecule type: protein

A;Residues: X', 27-28, X', 30-34, XX', 37 <AND2>

A;Note: confirmed presence of normal signal and absence of lipoprotein attachment

C;Genetics:

A;Gene: phoS1; psts

A;Start codon: GTG

C;Superfamily: phosphate-repressible phosphate-binding protein

C;Keywords: surface antigen

F;1-25/Domain: signal sequence #status predicted <SIG>

F;26-374/Product: phosphate specific transporter S #status experimental <MAT>

Query Match 36.1%; Score 1831; DB 1; Length 374;

Best Local Similarity 95.2%; Pred. No. 1.9e-80;

Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 59 RIKLEVSFKMRPAOP-----RCGSKPPSGSPETGAGGTATTTPASSPVTLAETGSTALL 112

Db 2 KIRLHTLLAVTAAPLLLAAAGCGSKPPSGSPETGAGGTATTTPASSPVTLAETGSTALL 61

QY 113 YPLFNLMPGAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNIGASDAYLSEGDVAAHKGLM 172

Db 62 YPLFNLMPGAPHERYPNVTTTAQCTGSGAGIAQAAAGTNNIGASDAYLSEGDVAAHKGLM 121

QY 173 NIALAISAOQVNNYLPVSEHLKNGKVLAAVGGTITKTDWDDPOIALNPGVNLPGTAVV 232

Db 122 NIALAISAOQVNNYLPVSEHLKNGKVLAAVGGTITKTDWDDPOIALNPGVNLPGTAVV 181

QY 233 FLHRSDSGDTFLFTQVLSKQDPGKSGFPGTGTTFVFPVPGALGNGGMMVTGCAET 292

Db 182 FLHRSDSGDTFLFTQVLSKQDPGKSGFPGTGTTFVFPVPGALGNGGMMVTGCAET 241

QY 293 PGCVAYIGISFLDQASORGLEAGLGNSSGNFLFLPDAQSIQAAAAGPASKTPANQAISMI 352

Db 242 PGCVAYIGISFLDQASORGLEAGLGNSSGNFLFLPDAQSIQAAAAGPASKTPANQAISMI 301

QY 353 DGPAPDGYPIINYEYATVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 412

Db 302 DGPAPDGYPIINYEYATVNNRQKDAATAQTLQAFHLWAITDGNKASFLDQVHFQPLPPAV 361

QY 413 VKLSDALIATISS 425

Db 362 VKLSDALIATISS 374

RESULT 3

H70802

Hypochemical protein Rv3974 - Mycobacterium tuberculosis (strain H37Rv)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: H70802

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-100 <COL>

A;Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAA17966.1; PID:g2960222

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: Rv3974

Query Match 9.6%; Score 485; DB 2; Length 100;

Best Local Similarity 100.0%; Pred. No. 5.5e-17;

Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 AEMKTDATLAQAGNPERISGDLKTDQVSTAGSLQGWGGAAGTAAQAAVVRFOEA 485

Db 2 AEMKTDATLAQAGNPERISGDLKTDQVSTAGSLQGWGGAAGTAAQAAVVRFOEA 61

QY 486 ANKQKQLDEISTNIRQAGVQYSGRADEEQQAALSSQMGF 524

Db 62 ANKQKQLDEISTNIRQAGVQYSGRADEEQQAALSSQMGF 100

RESULT 4

AG2377

Phosphate-binding periplasmic protein of phosphate ABC transporter all4575 [imported] - N

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AG2377

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: AG2377
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-392 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BA876274.1; PID:gl7133711; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Gene: all4575
 C/Superfamily: phosphate-repressible phosphate-binding protein

Query Match	9.5%; Score 482.5; DB 2; Length 392;
Best local Similarity	33.7%; Pred. No. 4.1e-16;
Matches	124; Conservative 64; Mismatches 151; Indels 29; Gaps 12;
QY	75 CGSKPPSGSPETGAG-AGTVATTASSP-----VTLAETGTTLLYPLFNLAGPAFHE 125
Db	29 CGGQGGDNATQGGSGTKADATASSPAKLDLGNVSLTCAGASFPAPLYASHTDLNK 88
QY	126 RYPNVTTTAQGTSGGAGIAQAAACTVIGASDAYLSEGD-AAHKGLMNIALAISAOQVN 184
Db	89 KYPNLQINYSVSGSAGVEQFIQTGTVFGASDVAMKDEBIQKQGVLLLEPVTAGGIVLA 148
QY	185 YNLPGVSEHLKLNCKVLAAMYQGTIKTWDDPQIALALNPGVNLPGTAVVPLHRSQSGDTF 244
Db	149 YNLPGVSTE-LNLPRVYTDILLGKIKTWDAPEIKAAANPNVNLPSQPIFVVVRSQSGGTG 207
QY	245 LFTOYLSKQDPEGKSGPFGTTFDPAVPALGNGNGMGMTVCAETPGCVAVIGISFL 304
Db	208 VFTKHLAAVSPE-WKSKVGECKSVSWPV---GVGKGNEGVTAIKQTQGAIGIYEYCY- 262
QY	305 DQASQRLGHAQLNGSGNFIPLDAQSTQAAAGAFSKTPANQAISMIDGPADG---YP 361
Db	263 --AKQNMI SYATLENKAGFKVYNDESASQTIA--AIQLPENLRAFV---PDPDGDGSP 315
QY	362 IINYEYAI VNNRQKDAATAOTLQAFLLHWAITDGNKASFLDCVHCPQPLPPVAVVKLSDALIA 421
Db	316 IVPSFWINAYKNYFDPVKAKANEAAIEYALTEGOKIS--GELGYIPLDQAVVQ-KTATVA 372
QY	422 TISSAEMK 429
Db	373 DOI SPFYK 380

RESULT 5
G82595
ABC transporter phosphate binding protein XF2141 [imported] - Xylella fastidiosa (strain G82595)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: G82595
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82595
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <SIM>
A:Cross-references: GB:A5004028; GB:A5003849; NID:g9107266; PIDN:AAF84940.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank. June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohman, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaiga, J.P.; Krueger, J.E.; Kurama, E.E.; Laigret, C.; Machado, A.M.; Machado, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.; Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.; A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr.V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka, J.

A; Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Vertovski-Almeida, S.; Vettore, A.L.; Zaccari, R.

A; Reference number: A59328

A; Contents: annotation

C; Genetics:

A; Gene: XP2141

C; Superfamily: phosphate-repressible phosphate-binding protein

Query Match 9.1%; Score 463; DB 2; Length 364;
Best Local Similarity 32.8%; Pred. No. 3.2e-15;
Matches 116; Conservative 67; Mismatches 151; Indels 20; Gaps 10;

Qy 76 GKSPPSGSPETGAGAGTATTTPASS---PVTLAETSTLLYPLENLWGPFAFHRYPNVT 131
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 21 GCKFSNDNQSGVSDGNSTTPSPAEEQTSKVISGAGASFIYPLISQWSADYNAACTN-K 79
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Qy 132 ITAQTGSCAGIAQAAGATVNIGASDAYILSEGDMAAHKLMIALAISQQVVYNLPVS 191
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 80 INYQSISSGGGIQAIIKAATIDFGSSDKPLDSSL-TQAGLGQFPSPAIGGVVPVVLNDIE 138
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Qy 192 E-HIKLAKGVLAAMYCGTIKTWDDPQIALNPNGVNLPGTAIVPLHRSDGSGDTFLFTQYL 250
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Db 139 PGKLRLTGPELLADIPLGKISKWNDAIISANPGLHLPTDKINIVHRSDGSCTTFNFSNYL 198
 :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
251 SKDPDGHKVCDCGETTVDFPAVDALGNGNGCVMTGTCAFTPCGVAVTCTSTFDPSOR 310

Db 199 SKVSAE-WKQKVGEGETSVOM---PGVGKGKNEGVSYYQQIKSGTGYVELAY---ALQN 251

Qy 311 GLGBAQLCNSSGNFLLPDAQSIQAAA--AGFASKTPANCAISMIDPAPDGPYIINYEYA 368

Db 252 KMSYITALQNAAGWQVPSAESFAAASNADSNAKDFNLVITNATGEA--AWPITAINFI 309

Qy 369 IVNNRQKDAAFTAQLQAFLHWAITDGNKASFLDQVHFQBLPPAVVKLSDALIAT 422

Db 310 LMRKQTDDAAQRKATLDFFKWSFENGQKQA--NELHYVLPPLPNLVQIEAYWAS 361

RESULT 6
H72756
probable phosphate-binding periplasmic protein APE0045 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
R:Accession: H72756
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ku
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72756
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <KAW>
A:Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA78954.1; PID:G5103433
A:Experimental source: strain K1

[illegible]

	Matches	108;	Conservative	69;	Mismatches	152;	Indels	14;	Gaps	9;
Qy	75	CGSKPPSG-	GPETGAGATGTAATT--PASSPVTLAETGTLLYPLFNLWGFAPHERYENVT	131	:	:	:	:	:	:
Dd	30	CGGGGGGD	TQAQTGGGDAITTTAADAFASKVSLTGAGASFPAPLYQGWFALNQAVERNLE	89	:	:	:	:	:	:
Qy	132	ITAQGTCSGAGIA	AAAAAGTVNIGASDAYLSGDMAAHKG-LMNTALAISAAQQVNYNLPGV	190	:	:	:	:	:	:
Dd	90	VNYQSVCSGAGVEQFMSKTYDVF	GSADVAMDDBEIAKNGEVVMLPMTAGSIWMAYINLPGV	149	:	:	:	:	:	:
Qy	191	SEHLKUNGKVLA	AAMYQGTITKTWDPTIALNPNGVNLPGTAVVLPHRSDDSGSDTFLFTQYL	250	:	:	:	:	:	:
Dd	150	-EGLKSQEALAGIMLGNI	KWNDPKLIVADNPDLTLPRPIVTVVHRSDDSGSTTAVFTMNIL	208	:	:	:	:	:	:
Qy	251	SKODPEGWGXSPGGTGTVDPAVPDGA-	LGENGCGMWTCGAETPCGCVAYICGISFLDOAQSG	309	:	:	:	:	:	:
Dd	209	AAMSPE-PKETIGDGKTVEMPTSXGKF	FGGKNEGEGVTAGIQONEGAICYVEYG- --ATN	264	:	:	:	:	:	:
Qy	310	RGLGEAQLGNSSGNFLLPDAQSIOAAA	AGAASFASKTPANCAISMIDGPADGVPITINYBVI	369	:	:	:	:	:	:
Dd	265	NNLTMASLNQKDGFVVPDTDENASATLIA-	AVEUPENIREBITPNAGASDSIPVITYTNWL	322	:	:	:	:	:	:
Qy	370	VNNRQKAATAQTLQALFWHAITDGNKASFLDQVHFQFLPPAV	412	:	:	:	:	:	:	:
Dd	323	LYPOYAADAERAKGIEAMVEFGINEGO-	-TWAPTILGYVPLPONV	363	:	:	:	:	:	:

RESULT 10
AI0499

probable phosphate-binding periplasmic protein pstS [imported] - *Yersinia pestis* (strain C/Species: *Yersinia pestis*
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C/Accession: A10499
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: A10499
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-346 <KUR>
A/Cross-references: GB:AL590842; PIDN:CAC93566.1; PID:gl5982006; GSPDB:GNO0175
C/Genetics:
A/Gen: pstS

A; gene: *pscs*
C; Superfamily: phosphate-repressible phosphate-binding protein

Query Match	8.7%	Score 440;	DB 2;	Length 345;
Best Local Similarity	33.4%	Pred. No. 3.7e-14;		
Matches 110;	Conservative	60;	Mismatches 141;	Indels 18;
Gaps	10;			

Qy	90	AGIVATTPAS--SPVTLAETGSTLLYPFLNUGFAFHERYENVTITAQGTSGSAGIAQAA	147
Db	13	AANTISMTAVSAFAAASLTGAGATPPAPVAKWADSYQKETCN-KINYGGIGSSGGVKKII	71
Qy	148	AGTVNIGASDAYLSEGDMAAHKGLMINTALISAQQVNNYLPGV-SEHLKMGKVLAAMYQ	206
Db	72	ANTVDFGASDAPLTDEKLAT-EGLFQPTVIGGVVLAVINPQIKSGELTLDGKTLDGIYL	130
Qy	207	GTIKTWDDPQTAALNPGVNLPGTAVVPLHRSDGSDTFLFYQYLSKQDPBEGWKGSPGCT	266
Db	131	GTVKKWNDDPALVKLNPGVKLPQONIAVVRADGGTSEVFTSYLAKVNAB-WKEKVGAGS	189
Qy	267	TVDPPAVPGALGENGGWVFGCAETPGCVAYIGISFLDQASQGLGEAQIUNGSSGNFLL	326
Db	190	TVNNPT--GLGGKNGDCAAFVQRLPGSIGVVEYAY--AKONNIAVTKLISADGRPVS	243
Qy	327	PDAGSIQAAAGGFA-SKTPANQAISMTIDGPADGPYPLINYEYAIVNNRQKDAATAQTLOA	385
Db	244	PTEHSFSAAGVDMWSKFAQ--DLTNQKGDVDWVPTSTTFLVHKQEOKNAANGTEVLK	300
Qy	386	FLHWAITDGNKASFELDQVHFQPLPFAVVK	414

```

Db      301 FFDGWTGHAKQA--NELDYATLPAEYVE 327
| | | | | : : : : | | | :
RESULT 11
H91211 phosphate-binding periplasmic protein PstS [imported] - Escherichia coli (strain
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: H91211
C/Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.;
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, S.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: H91211
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-346 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA838087.1; PID:G13364139; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain RIMD 050952
C/Genetics:
A/Gene: Ecs4664
C/Superfamily: phosphate-repressible phosphate-binding protein

```

Query Match	8.3%	Score 419;	DB 2;	Length 346;	
Best local Similarity	32.5%;	Pred. No. 3.7e-13;			
Matches 109;	Conservative	56;	Mismatches 144;	Indels 26;	Gaps 10;
QY	92	TVATTTPAS-----SPVTLAETGSTALLPLENLWGPAFHERYPNVVTITAGTCSGA	141		
DB	7	TVATVTAATLSMGAFSVFASITGAGATPPAPVYAKWADTYKETGN-KVNYGIGSSG	65		
QY	142	GIAQAAAGTVNIGASDAYLSGEGVAAKHGLMNLALISAQVNVNLPGV-SEHLKLGKV	200		
DB	66	GVQIIANTVDFGASDAPLSD-EKLAQEGLFQPPVIGGVVLAVINFLKSGELVLDGKT	124		
QY	201	LAAMYQTIKTDDPQIAALNPGVNLPGTAVVPLHRSDGSGDTPLFQYVSKQDPBGWK	260		
DB	125	LGBIYLKIKRWDEATAKLNPLGLPSONIAVVVRADGSGTSVFVTSYLAKNEE-WKN	183		
QY	261	SPGFGTTVDPAVPGALGENGGWVGTCAETPGCVAYIGISFLDQASORGLGEAQLCNS	320		
DB	184	NVGTGSTVKWPI---GUGGKNGDIIAFAVORLPGAIGYVEYAV---AKNNLVATKLISA	237		
QY	321	SGNFLPDAQSIQAAAGFA-SKTPANQAIASMIDGPAPDGPYPIINYSYAIVNNRQKDAAT	379		
DB	238	DGKPEVLPTSENFANAAGKADWSKTFQAQ---DLTNQKGEDAWPITSTFFILHKKDQKPEQ	294		
QY	380	AQTLQAFELHWAITDGNKASFLDQVHFOPLPPAVVK	414		
DB	295	GTEVLKFFDWAYKTGAQQA---NLDLDVASLPDSVVE	327		

RESULT 12
H86057 phosphate-binding periplasmic protein PstS [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: H86057
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.;
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.A.;
Nature 409, 529-533, 2001
A:Title: Genomic sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H86057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:AE005174; NID:gl2518575; PIDN:AGS8924.1; GSPDB:GN00145
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: pstS

C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 8.3%; Score 419; DB 2; Length 346;
 Best Local Similarity 32.5%; Pred. No. 3.7e-13;
 Matches 109; Conservative 56; Mismatches 144; Indels 26; Gaps 10;
 QY 92 TVATTPAS-----SPVTLAETGSTALLYPLFNLMGPAHERYVNVITTAQGTGSGA 141
 DB 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYQKTCGN-KVNYQGIGSSG 65
 QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKGLMNTALISAQQVNNYLPV-SEHLKNGKV 200
 DB 66 GVQKIIANTVDFGASDAPLSD-EKLAQEGLFQFPPTVIGGVVLAVNIPGKSGELVLDGKT 124
 QY 201 LAAMYQGTITKWDPPQIAALNPGVNLPGTAVVPLHRSDSGDTFLFTQYLSKQDPGSGWK 260
 DB 125 LGDIYLGKIKKWDDEATAKLNPKLPSONIAVVRADSGTSFVFTSYLAKVNEE-WKN 183
 QY 261 SPFGTGVDFPVPALGNGGNGMTGCAETPCGVAYIGISFLDOASQRLGEAQLGNS 320
 DB 184 NVGTGSTVKWPI---GLGKGNDAIAFVORLPGAIGYVEYAY---AKNNLAYTKLISA 237
 QY 321 SGNELLFDAQSIQAAAAGFA-SKTPANQAISMDGPAPOGYPPIINVEYAVVNNRQKDAAT 379
 DB 238 DGKPVLPTEENFANAAGADWSKTFQA---DLTNQKGEDAWPITSTFILIHKDKKPEQ 294
 QY 380 AQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVK 414
 DB 295 GTEVLKFFDWAYKTGAQQA--NDLDYASLPDSVVE 327

RESULT 13

BYECPR

phosphate-repressible phosphate-binding protein precursor. [validated] - Escherichia coli

C;Species: Escherichia coli
 C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 01-Mar-2002
 C;Accession: A30277; A65176
 R;Surin, B.P.; Jans, D.A.; Fimmel, A.L.; Shaw, D.C.; Cox, G.B.; Rosenberg, H.
 J. Bacteriol. 157, 772-778, 1984
 A;Title: Structural gene for the phosphate-repressible phosphate-binding protein of Esch.
 A;Reference number: A30277; MUID:84135579; PMID:6321434
 A;Accession: A30277
 A;Molecule type: DNA
 A;Residues: 1-346 <SUR>
 A;Cross-references: GB:K01992; NID:G147255; PIDN:AAA24378.1; PID:G147256
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of Escherichia coli K-12.
 A;Reference number: A64720; MUID:97426617; PMID:9278503
 A;Accession: A65176
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-346 <BLAT>
 A;Cross-references: GB:AE000449; GB:U00096; NID:G2367269; PIDN:AACT67651.1; PID:G2367271;
 A;Experimental source: strain K-12, substrain MG1655
 R;Wang, Z.; Luecke, H.; Quioco, F.A.
 submitted to the Brookhaven Protein Data Bank, August 1996
 A;Reference number: A68487; PDB:1IXH
 A;Contents: annotation; X-ray crystallography, 0.98 angstroms, residues 26-346
 R;Yao, N.; Choudhary, A.; Ledvina, P.S.; Quioco, F.A.
 submitted to the Brookhaven Protein Data Bank, November 1995
 A;Reference number: A67875; PDB:2ABH
 A;Contents: annotation; X-ray crystallography, 1.70 angstroms, residues 26-346
 R;Luecke, H.; Quioco, F.A.
 submitted to the Brookhaven Protein Data Bank, April 1992
 A;Reference number: A51014; PDB:1ABH
 A;Contents: annotation; X-ray crystallography, 1.7 angstroms, residues 26-110, 'G', 112-20
 R;Luecke, H.; Quioco, F.A.
 Nature 347, 402-406, 1990
 A;Title: High specificity of a phosphate transport protein determined by hydrogen bonds.
 A;Reference number: A30645; MUID:91015319; PMID:2215649
 A;Contents: annotation; X-ray crystallography, 1.7 angstroms

C;Comment: This protein is a component of the inducible, high-affinity, phosphate-specific
 C;Genetics:

A;Gene: pstS; phoS
 A;Map position: 84 min
 C;Superfamily: phosphate-repressible phosphate-binding protein
 C;Keywords: phosphate transport
 F;1-35/Domain: signal sequence #status predicted <SIG>
 F;26-346/Product: phosphate-repressible phosphate-binding protein #status experimental <N
 F;35,63,81,160,164,166/Binding site: phosphate (Thr, Ser, Asp, Arg, Ser, Thr) #status ext

Query Match 8.1%; Score 413; DB 1; Length 346;
 Best Local Similarity 32.2%; Pred. No. 7.2e-13;
 Matches 108; Conservative 56; Mismatches 145; Indels 26; Gaps 10;

QY 92 TVATTPAS-----SPVTLAETGSTALLYPLFNLMGPAHERYVNVITTAQGTGSGA 141
 DB 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYQKTCGN-KVNYQGIGSSG 65
 QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKGLMNTALISAQQVNNYLPV-SEHLKNGKV 200
 DB 66 GVQKIIANTVDFGASDAPLSD-EKLAQEGLFQFPPTVIGGVVLAVNIPGKSGELVLDGKT 124
 QY 201 LAAMYQGTITKWDPPQIAALNPGVNLPGTAVVPLHRSDSGDTFLFTQYLSKQDPGSGWK 260
 DB 125 LGDIYLGKIKKWDDEATAKLNPKLPSONIAVVRADSGTSFVFTSYLAKVNEE-WKN 183
 QY 261 SPFGTGVDFPVPALGNGGNGMTGCAETPCGVAYIGISFLDOASQRLGEAQLGNS 320
 DB 184 NVGTGSTVKWPI---GLGKGNDAIAFVORLPGAIGYVEYAY---AKNNLAYTKLISA 237
 QY 321 SGNELLFDAQSIQAAAAGFA-SKTPANQAISMDGPAPOGYPPIINVEYAVVNNRQKDAAT 379
 DB 238 DGKPVLPTEENFANAAGADWSKTFQA---DLTNQKGEDAWPITSTFILIHKDKKPEQ 294
 QY 380 AQTLQAFHLWAITDGNKASFLDQVHFQPLPPAVVK 414
 DB 295 GTEVLKFFDWAYKTGAQQA--NDLDYASLPDSVVE 327

RESULT 14

AB0956

periplasmic phosphate-binding protein [imported] - Salmonella enterica subsp. enterica se
 C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AB0956
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AB0956
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-346 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD03142.1; PID:G16504777; GSPDB:GN00176
 C;Genetics:

C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 8.1%; Score 411; DB 2; Length 346;
 Best Local Similarity 32.6%; Pred. No. 9e-13;
 Matches 109; Conservative 52; Mismatches 147; Indels 26; Gaps 10;
 QY 92 TVATTPAS-----SPVTLAETGSTALLYPLFNLMGPAHERYVNVITTAQGTGSGA 141
 DB 7 TVATVVAATLSMSAFSAFASLTGAGATPPAPVYAKWADTYQKTCGN-KVNYQGIGSSG 65
 QY 142 GIAQAAAGTVNIGASDAYLSEGDVAAHKGLMNTALISAQQVNNYLPV-SEHLKNGKV 200
 DB 66 GVQKIIANTVDFGASDAPLSD-EKLAQEGLFQFPPTVIGGVVLAVNIPGKSGELVLDGKT 124

RESULT 15

Query Match · 8.0%; Score 406; DB 2; Length 383;

Best Local Similarity 29.0%; Pred. No. 1.8e-12;
Matches 113; Conservative 64; Mismatches 149; Indels 64; Gaps 13;

```

QY      130  VTTAQTGTGAGIAQAAGTWNIGASD-----AYLSEGDWAHKGIMNIALAISA 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      77  VKIEYEGGGSGYGOFAELKGLTIDGRDPPVKESTWKKFLQTGP-----LQPEIVAGA 131

```

181 QQVNNYIPGVSEHLKNGKVLAAVYQGTIKTDWDPQIAALNPGVNLPGTAVVPLHRSDGS 240
132 WAAAAVPEVNF-LKTAREVIAKTFIGETVYMDQDPAIKRLNPNKTI PHKKTIVTHRSDAS 190

241 GDTEFLTQYLKQDPHGKSGPGTGTVDPA--VPGALGNGNGMVTGCAETPGCVAY 298
191 GATTAETPTVYSLIATSWKAPVYGVKGVKMDVMDVYGCCTACAGKGVKVAALKVMTSTAY 249

299 IGISFLDOASORGIGEAOLGNSSGNFLLPDAQSTCAAAAGFASKTP-----ANQAISMI 352

Qy 353 DGPADGVPINYEYIV-----NNRQKDAATAQILOAFLHWAITGKNKASFLDQVHQPL 408

DB 409 PPVVKUSDALIATISSAEMKTDATLAQE 438
307 DAPGENSYPIVAF:THLLVWQNGGRKHYPFKAKATKDFKWLITGGQAP-----ENL 339

DB 359 AFGVGLPREV-----ABIGKAVEMLEE 382

Job time : 21.507 secs

This Page Blank (uspio)

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AY029285; AAK1576.1; -
 DR EMBL; AL022120; CAAL1793.1; -
 DR EMBL; AE007191; AAK48364.1; -
 DR PIR; G70803; G70803.
 DR TIGR; MT3996; -
 DR TubercuList; Rv3881c; -

KW Antigen; Complete proteome.
 FT DOMAIN 306 428 GLY-RICH.
 SQ SEQUENCE 460 AA; 47593 MW; 7528743226AD7A71 CRC64;

Query Match 46.8%; Score 2375; DB 1; Length 460;
 Best Local Similarity 100.0%; Pred. No. 1.1e-98;
 Matches 459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 TQQTVDQOEILNRANEVEAPMADPTDPTTPELTAANKAAQOLVLSADNNREYLA 584
 DB 2 TQQTVDQOEILNRANEVEAPMADPTDPTTPELTAANKAAQOLVLSADNNREYLA 61
 QY 585 AGAKERORLATSURNAKAYGEVDEEAATLNDGEGTVOAESAGAVGDSAEALTDTPR 644
 DB 62 AGAKERORLATSURNAKAYGEVDEEAATLNDGEGTVOAESAGAVGDSAEALTDTPR 121
 QY 645 VATAGEPFDLKEAARKLTGQOGASLAHFADGWNFTNLTLQGDVKKFRFGDNWEGDAA 704
 DB 122 VATAGEPFDLKEAARKLTGQOGASLAHFADGWNFTNLTLQGDVKKFRFGDNWEGDAA 181
 QY 705 TACEASLDQOROLHMAKLSAMAKAQYVAQLHWARREPTYEDIVGLRELVAENPS 764
 DB 182 TACEASLDQOROLHMAKLSAMAKAQYVAQLHWARREPTYEDIVGLRELVAENPS 241
 QY 765 ARQILPVAYEYQORSEKVLTEYNKAALEPVNPPPPPAIKIDPPPPQEOGLIPGFLM 824
 DB 242 ARQILPVAYEYQORSEKVLTEYNKAALEPVNPPPPPAIKIDPPPPQEOGLIPGFLM 301
 QY 825 PPSDGSVPTGTMPPAPVPTPTGSGGGLPAPTAQAQLTSAGREAAALSGDVAVKAAASLG 884
 DB 302 PPSDGSVPTGTMPPAPVPTPTGSGGGLPAPTAQAQLTSAGREAAALSGDVAVKAAASLG 361
 QY 885 GGGGGVPSAPLGSATGGAESVAPAGAGDIAGLQGRAGGAGALGGGGMGMPGAAHQOQ 944
 DB 362 GGGGGVPSAPLGSATGGAESVAPAGAGDIAGLQGRAGGAGALGGGGMGMPGAAHQOQ 421
 QY 945 GGAKSGSQOEDBALYTEDRAWTEAVIGNRRRDSKESK 983
 DB 422 GGAKSGSQOEDBALYTEDRAWTEAVIGNRRRDSKESK 460

RESULT 2
 PSTL_MYCTU STANDARD; PRT; 374 AA.
 AC P15712; O05868;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding protein 1 precursor (PBP-1) (Pbts-1) (Protein
 DE antigen B) (PAB) (Antigen Ag78).
 GN PSTL OR PHOSI OR RV0934 OR MT0961 OR MTCV0809.05C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89307568; PubMed=2545626;
 RA Andersen A.B.; Hansen E.B.;
 RT "Structure and mapping of antigenic domains of protein antigen b, a
 RT 38,000-molecular-weight protein of Mycobacterium tuberculosis";
 RL Infect. Immun. 57:2481-2488(1989).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=H378V;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T.; Brosch R.; Parkhill J.; Garnier T.; Churcher C.; Harris D.,
 RA Gordon S.V.; Eiglmeier K.; Gas S.; Barry C.E. III; Tekala F.,
 RA Badcock K.; Basham D.; Brown D.; Chillingworth T.; Connor R.,
 RA Hovnsby T.; Jagels K.; Krogh A.; McLean J.; Moule S.; Murphy L.,
 RA Oliver S.; Osborne J.; Quail M.A.; Rajandream M.A.; Rogers J.,
 RA Rutter S.; Seeger K.; Skelton S.; Squares S.; Squares R.,
 RA Sulston J.E.; Taylor K.; Whitehead S.; Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RN Nature 393:537-544(1998).
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D.; Alland D.; Eisen J.A.; Carpenter L.; White O.,
 RA Peterson J.; DeBoy R.; Dodson R.; Gwinn M.L.; Haft D.; Hickey E.,
 RA Kolonay J.F.; Nelson W.C.; Umayam L.A.; Ermolaeva M.D.; Salzberg S.L.,
 RA Delcher A.; Utterback T.; Weidman J.; Khouri H.; Gill J.; Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Required for binding-protein-mediated phosphate
 CC transport.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Probable).
 CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL curation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M30046; AAA25374.1; -
 DR EMBL; Z95209; CAB0484.1; -
 DR EMBL; AE006981; AAK45208.1; -
 DR PIR; F70584; F70584.
 DR HSSP; P06128; 11XH.
 DR TIGR; MT0961; -
 DR TubercuList; RV0934; -
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
 KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KW Antigen; Complete proteome.
 FT SIGNAL 1 23 POTENTIAL..
 FT CHAIN 24 374 PHOSPHATE-BINDING PROTEIN 1..
 FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 374 AA; 38243 MW; 6334968191FF38AA CRC64;

Query Match 36.1%; Score 1831; DB 1; Length 374;
 Best Local Similarity 95.2%; Pred. No. 1.1e-74;
 Matches 355; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
 QY 59 RIKLEVSFKMRPAQP-----RCGSKPPSPETGAGAGTVAITPASPVTLAETGSLTLL 112
 DB 2 KIRLHTLLAVLTAAPLLAAAGCGSKPPSPETGAGAGTVAITPASPVTLAETGSLTLL 61
 QY 113 YPLENLWGPAPHERYPNVTTITAGTSGAGIAQAAGATVNVIGASDAYLSGDDAAHKGML 172
 DB 62 YPLENLWGPAPHERYPNVTTITAGTSGAGIAQAAGATVNVIGASDAYLSGDDAAHKGML 121
 QY 173 NIALAISAQVQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIALNPGVNLPGTAVV 232
 DB 122 NIALAISAQVQVYNLPGVSEHLKNGKVLAAAYQGTIKTWDDPQIALNPGVNLPGTAVV 181
 QY 233 PLHSDSGSGDTFLFTQVLSKQDPGKWSKSPFGTITVDFFPAVPGALGNGNGMVTGCAET 292

Db 182 PLHRSQSGDTLFTQYLSKQDPGKSGFGFTTVDPAVPGALGNGNGMVGCAET 241
QY 293 PCGVAYIGISFSDQASQRLGSAQNLGNSGNFLLPDAQSIQAAAAGFASKTPANQAIWI 352
Db 242 PCGVAYIGISFSDQASQRLGSAQNLGNSGNFLLPDAQSIQAAAAGFASKTPANQAIWI 301
QY 353 DGPADPGYPIINYEVAIVNNRQKDAATQTLQAFILHWAITDGNKASFIDQVHFQPLPPAV 412
Db 302 DGPADPGYPIINYEVAIVNNRQKDAATQTLQAFILHWAITDGNKASFIDQVHFQPLPPAV 361
QY 413 VKLSDALIATISS 425
Db 362 VKLSDALIATISS 374

RESULT 3

CF10 MYCTU STANDARD; PRT; 99 AA.
AC 069739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 10 kDa culture filtrate antigen cfp10 (Secreted antigenic protein
DE MTA-10).
GN CFP10 OR LHP OR WTA10 OR RV3874 OR WT3988 OR MTV027.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC STRAIN=H37Rv;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
RT low-molecular-mass culture filtrate protein (Cfp-10).";
RL Microbiology 144:3195-3203 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies A., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Singh B., Siddiqui Z., Singh S., Sharma P.;
RT "RV3874 (msa-10) gene of a clinical isolate of Mycobacterium
RT tuberculosis from India.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; AF004671; AAC83445.1; -
DR EMBL; AL022120; CAAL7966.1; -
DR EMBL; AE007190; AAK48356.1; -
DR EMBL; AF419854; AAL14999.1; -
DR PIR; H70802; H70802.
DR TIGR; MT3988; -
DR Tuberculist; RV3874; -
KW Antigen; Complete proteome.
FT INIT MET 0
SQ SEQUENCE 99 AA; 10663 MW; ESCAE6A996C5489D CRC64;
Query Match 9.6%; Score 485; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 9.7e-16;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 426 AEMKTDATLAQAGNFERISGLKTKQIDQVESTAGSLQGWGAGTAAQAAVVRQEA 485
Db 1 AEMKTDATLAQAGNFERISGLKTKQIDQVESTAGSLQGWGAGTAAQAAVVRQEA 60
QY 486 ANKQKQELDEISTNIRQAGVQYGRADDEQQQALSSQMGF 524
Db 61 ANKQKQELDEISTNIRQAGVQYGRADDEQQQALSSQMGF 99
RESULT 4
PSTS XYLYFA
ID PSTS XYLYFA STANDARD; PRT; 364 AA.
AC Q9PBK3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR XP2141.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach P.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferri J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Honeiss J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kurana E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.P., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

```
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
CC TRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: AE004028; AAF84940.1; -.
DR HSP: G82595; G82595.
DR InterPro: IPR006059; SBP_bac_1.
DR Pfam: PF01547; SBP_bac_1; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 364 AA; 38736 MW; 8609CFAA159D4277 CRC64;

Query Match 9.1%; Score 463; DB 1; Length 364;
Best Local Similarity 32.8%; Pred. No. 4.1e-14;
Matches 116; Conservative 67; Mismatches 151; Indels 20; Gaps 10;

QY 76 GSRPPSPGPGAGATVATPASS-----PVTAETGSLTLLYPLFNLWGFAPHERYPNV 131
DB 21 GCRPSNDQSTGVSDGNSSTPSSAEQTSKVKISGAGAFYPLISQWSADYNAATGN-K 79

QY 132 ITAGCTSGAGIAQAAGTNYIGASDAYLSEGDMAAHKLMLNLAISAQVNNYLPVGS 191
DB 80 INQISGSGGIAQIKATIDFGSSDKPLDSEL-TQALGQFSPSALGGVVPVNLDNTE 138

QY 192 E-HKLNKGVLAAMYQTIKTDDPQIAALNPGVNLPGTAVVPLHRSDGSDTFLTQYL 250
DB 139 PGKRLATGCLPLADIFLKGISKNDAATISAANPGLHLPDPTKINTVHRSDGSGTTFNFSNYL 198

QY 251 SKQDPEGWKGSPGFTTVPVAVPGALGNGNGWGTGCAETPGCVAYIGISFLDQASOR 310
DB 199 SKVSAE-WKQKVGEGTSVQW---PGVGKGNGEVSAYVQCIKSGYVELAY---ALQN 251

QY 311 GLGEAQIGNSGNNFLPDPAQISQAAA--AGFASKTPANQAISMIDGPADGVPPIIYNEVA 368
DB 252 KMSYTAQLNAAQGVQPSAESFAAAAGNADWSNAKDFNLVITNATGEA--AWPITATNFI 309

QY 369 IVNRRQKDAATAQLQAFLEHWAIDGNKASFLDQVHFQPLPRAVVKLSDALIAT 422
DB 310 LMRQTKDAARQKATLDFKWSFENGQKQA--NELHYVPLPVLNIVQIEAYWAS 361

RESULT 5
PSTS_RHILO
ID PSTS_RHILO STANDARD; PRT; 351 AA.
AC Q98FL2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR ML13723
OS Rhizobium loti (Mesorhizobium loti)
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1],
```

```
RP SEQUENCE FROM N.A.
RC STRAIN:MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Katanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA RES. 7:331-338(2000).
CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
CC TRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: AP003002; BAB50555.1; -.
DR HSP: P06128; 11XH.
DR InterPro: IPR006059; SBP_bac_1.
DR Pfam: PF01547; SBP_bac_1; 1.
KW Phosphate transport; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 351 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 351 AA; 36766 MW; 2710D1BC0417574D CRC64;

Query Match 8.8%; Score 444; DB 1; Length 351;
Best Local Similarity 34.4%; Pred. No. 2.7e-13;
Matches 112; Conservative 51; Mismatches 149; Indels 14; Gaps 8;

QY 95 TTPASSPVTLAETGSLTLLYPLFNLWGFAPHERYPNVITAGCTSGAGIAQAAGTNYIG 154
DB 21 TLSAAIAADLSAGSTPIYPVFAKWADTY-KKDTGVGLNYQSIGSGGIGKQVIKTTFG 79

QY 155 ASDAYLSEGDMAAHKLMLNLAISAQVNNYLPVSE-HKLNKGVLAAMYQTIKTWD 213
DB 80 ATDKFMSDADLEKX-GLVQPFPMVNGGIVPIVNLTFGKPELVDLGDGKTLAQIYLGAITTD 138

QY 214 DPQIAALNPGVNLPGTAVVPLHRSDGSDTFLTQYLSKQDPEGWKGSPGFTTVPVAV 273
DB 139 DAAIKALNPSLTLPSTALVAVHRSDGSGTTFNINLVKLSFD-WKDKVGSDDTAVENTP- 196

QY 274 PGALGNGNGWGTGCAETPGCVAYIGISFLDQASQRLGEAQIGNSGNNFLPDPAQISQ 333
DB 197 --GVGAKGSEGVANTVKTDGGIGYVEYAY---AKQNNLSYSKMLNAAAGKVVEPSLSEFG 251

QY 334 AAAAGFASKTPANQAISMIDGPADGVPPIIYNEVAIVNRRQKDAATAQLQAFLEHWAID 393
DB 252 AAASNADPKGAKNFNVIITNEPGDTTWFIAASTWLIHKAPDDDAATGEALKFFAWAYKD 311

QY 394 GNK-ASFLDQVHFQPLPRAVVKLSDA 418
DB 312 GKETAALDYV--SIPDSVVDLIKA 334

RESULT 6
PSTS_ECOLI
ID PSTS_ECOLI STANDARD; PRT; 346 AA.
AC P06128; P76744;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding periplasmic protein precursor (PBP).
GN PSTS OR PHOS OR B3728 OR SF3727.
OS Escherichia coli, and
```

OS *Shigella flexneri*
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 623;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC SPECIES=E.coli;
 RX MEDLINE=84135579; PubMed=6321434;
 RA Surin B.P., Jans D.A., Fimmel A.L., Shaw D.C., Cox G.B., Rosenberg H.;
 RT "Structural gene for the phosphate-repressible phosphate-binding
 RT protein of *Escherichia coli* has its own promoter: complete nucleotide
 RT sequence of the phos gene.";
 RL J. Bacteriol. 157:772-778(1984).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC SPECIES=E.coli;
 RX MEDLINE=84135599; PubMed=6365894;
 RA Magota K., Otsuji N., Miki T., Horiuchi T., Tsunasawa S., Kondo J.,
 RA Sakiyama F., Amemura M., Morita T., Shinagawa H., Nakata A.;
 RT "Nucleotide sequence of the phos gene, the structural gene for the
 RT phosphate-binding protein of *Escherichia coli*.";
 RL J. Bacteriol. 157:903-917(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;
 RX MEDLINE=93315143; PubMed=7686882;
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*
 RT genome: organizational symmetry around the origin of replication.";
 RL Genomics 16:551-561(1993).
 RN [4]
 RP SEQUENCE OF 26-37.
 RC SPECIES=E.coli; STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of *Escherichia coli* K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=2227406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 RT through comparison with genomes of *Escherichia coli* K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RC SPECIES=E.coli;
 RX MEDLINE=91015319; PubMed=2215649;
 RA Luecke H., Quirocho F.A.;
 RT "High specificity of a phosphate transport protein determined by
 RT hydrogen bonds.";
 RL Nature 347:402-406(1990).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.89 ANGSTROMS) OF MUTANT ASN-81.
 RC SPECIES=E.coli;
 RX MEDLINE=97372887; PubMed=9228942;
 RA Wang Z., Luecke H., Yao N., Quirocho F.A.;
 RT "A low energy short hydrogen bond in very high resolution structures
 RT of protein receptor-phosphate complexes.";
 RL Nat. Struct. Biol. 4:519-522(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RC SPECIES=E.coli;
 RX MEDLINE=98337779; PubMed=9671506;
 RA Hirschberg M., Henrick K., Haire L.L., Vasisth N., Brune M.,
 RA Corrie J.E., Webb M.R.;
 RT "Crystal structure of phosphate binding protein labeled with a

RT
 RL Biochemistry 37:10381-10385(1998).
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN
 CC -MEDIATED PHOSPHATE TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY PHOSPHATE DEPRIVATION AND SUBJECT TO POSITIVE
 CC CONTROL BY THE PHO B GENE PRODUCT AND TO NEGATIVE CONTROL BY THE
 CC PHO R GENE PRODUCT.
 CC -1- SIMILARITY: BELONGS TO THE PSTS FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; K01992; AAA24378.1; -;
 DR EMBL; L10328; AAA62079.1; -;
 DR EMBL; AE000449; AAC76751.1; -;
 DR EMBL; AE015380; AAN45173.1; -;
 DR PIR; A30277; BYECPR.
 DR PDB; 2ABH; 21-APR-97.
 DR PDB; 1OIB; 21-APR-97.
 DR PDB; 1PBP; 15-OCT-94.
 DR PDB; 1QU1; 11-JUL-96.
 DR PDB; 1QU3; 11-JUL-96.
 DR PDB; 1QUK; 11-JUL-96.
 DR PDB; 1QUL; 11-JUL-96.
 DR PDB; 1IXG; 04-FEB-98.
 DR PDB; 1IXH; 04-FEB-98.
 DR PDB; 1IXI; 22-OCT-97.
 DR PDB; 1A40; 23-MAR-99.
 DR PDB; 1A54; 14-OCT-98.
 DR PDB; 1A55; 14-OCT-98.
 DR SWISS-2DPAGE; P06128; COLI.
 DR EcoGene; EG10734; pstS.
 DR InterPro; IPR005673; Peri-phosph.
 DR InterPro; IPR006059; SSB_bac_1.
 DR Pfam; PF01547; SSB_bac_1_1.
 DR TIGRFAMs; TIGR00975; Ss0107803; 1.
 KW Phosphate transport; Transport; Periplasmic; Signal; 3D-structure;
 KW Complete proteome.
 FT SIGNAL 1 25
 FT CHAIN 26 346 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
 FT STRAND 28 33
 FT TURN 36 37
 FT HELIX 38 52
 FT STRAND 55 60
 FT HELIX 63 71
 FT TURN 72 73
 FT STRAND 77 80
 FT HELIX 86 92
 FT TURN 93 93
 FT STRAND 94 108
 FT TURN 111 112
 FT TURN 115 116
 FT STRAND 119 120
 FT HELIX 122 129
 FT TURN 130 131
 FT STRAND 135 135
 FT TURN 136 137
 FT HELIX 139 144
 FT TURN 146 147
 FT STRAND 154 154
 FT STRAND 156 160
 FT HELIX 165 177
 FT TURN 179 184
 FT TURN 185 185
 FT STRAND 197 199
 FT HELIX 201 211
 FT TURN 213 214

FT STRAND 215 220
 FT HELIX 221 225
 FT TURN 226 228
 FT STRAND 230 231
 FT STRAND 232 235
 FT TURN 233 238
 FT STRAND 237 238
 FT STRAND 241 242
 FT HELIX 246 252
 FT TURN 253 255
 FT TURN 258 260
 FT STRAND 267 267
 FT TURN 272 273
 FT STRAND 277 286
 FT STRAND 288 288
 FT HELIX 292 308
 FT HELIX 310 315
 FT TURN 316 317
 FT STRAND 319 320
 FT HELIX 323 336
 FT STRAND 338 338
 FT TURN 340 341
 FT STRAND 344 344
 SQ SEQUENCE 346 AA; 37024 MW; 867DA7199C2C87ED CRC64;

 Query Match 8.1%; Score 413; DB 1; Length 346;
 Best Local Similarity 22.2%; Pred. No. 6.4e-12;
 Matches 108; Conservative 56; Mismatches 145; Indels 26; Gaps 10;

 QY 92 TVATTPAS-----SPVTIAETGSTALLYPLNLPWPAFHRYPNVTITTAQSTGSGA 141
 DB 7 TVATVAAATLSMAFSVFAEASLTGAGATPPADVYAKWADTYKGTGN-KVNYQIGISG 65

 QY 142 GICAAAGTINIGASDAYLSEGDMAHKGMLNLTALISAQVNNYLPGV-SEHLKNGKV 200
 DB 66 GVKQITANTVDFGASDAPLSD-EKLAQEGLFQFPPTVIGGVVAVNIPGLKSELVLDGKT 124

 QY 201 LAAMYQOTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGDFTFLTYQLSKQDPGNGK 260
 DB 125 LGDIYLGKIKWDDAELAKNLGKLPSQNIIVRRADSGTSFVTSYLAKNEE-WKN 183

 QY 261 SPFGITVDPVAPVPGALGENGNGMVTGCAETPCGVAYIGISFLDQASQRLGEAQLGNS 320
 DB 184 NVGTGSTRVWPI---GLGGKXNDGIAAFVQRLPGAIGYVEYAY---AKQNNLAYTKLISA 237

 QY 321 SGNFFLPDAQSIOAAAAGPA-SKTPANQALSMIDGPDGPGPYPIINYEYAIYVNNRQKDAAT 379
 DB 238 DGRFVSSTEENFANAAGADWSKTFQAQ---DLTNQGEDAWPITSTFFILHKDQKKPEQ 294

 QY 380 AQTQLAFLHWAITDGNKASFLDQVHFQPLPPAVVK 414
 DB 295 GTEVLKFFDWAYKTGAQA--NDLDVASLPDSVVE 327

RESULT 7

ID PSTS PASMU STANDARD; PRT; 344 AA.
 AC QCN34;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PSTS OR PM0436.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittan T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc., Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -!- FUNCTION: REQUIRED FOR BINDING-PROTEIN-MEDIATED PHOSPHATE
 CC TRANSPORT (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE006079; AAK02520.1; ALT_INIT.
 DR HSSP; P06128; IIXH.
 DR Pfam; PF01547; SBP_bac_1; 1.
 KW Phosphate transport; Transport; Periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 344 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
 SQ SEQUENCE 344 AA; 37362 MW; 6378650112CF9161 CRC64;

 Query Match 8.1%; Score 411.5; DB 1; Length 344;
 Best Local Similarity 31.7%; Pred. No. 7.4e-12;
 Matches 108; Conservative 59; Mismatches 159; Indels 15; Gaps 8;

 QY 92 TVATTPASSPVTLAETGSTALLYPLNLPWPAFHRYPNVTITTAQSTGSGAGIAQAAGTV 151
 DB 14 TFAUTTOAQITIGA--GASFPYPIYAKWA-SMVEKQTKGNKNYQISGGGQQIITAKTI 70

 QY 152 NIGASDAYLSEGDMAHKGMLNLTALISAQVNNYLPVSE-HLKLNGKVLAAMYQOTIK 210
 DB 71 DFGASDPMKAEALQAQ-LIQPFAIIGTVPVVNLPEITAGQLKLSGEVLADIFLGKIK 129

 QY 211 TWDDPQIAALNPGVNLPGTAVVPLHRSDSGDFTFLTYQLSKQDPGNGKSPGFTVDF 270
 DB 130 KWNDFAIKLNQGANLPDKAIIVHRSDDSGTFTGNTNLSKVSTE-WKETVGGKSVKW 188

 QY 271 PAVPGALGENGNGMVTGCAETPCGVAYIGISFLDQASQRLGEAQLGNSGNFLPLDQAQ 330
 DB 189 PTGGGKNGEVAAYVSKIKYSGYVEYA-----YAKQNLAWASLONKAGQVQPSAE 242

 QY 331 SIQAAAAGFASKTPANQALSMIDGPDGPGPYPIINYEYAIYVNNRQKDAATQTTQAFLHWA 390
 DB 243 SFMAAANAQWESAVGMVILTNEEGDTSWPTVAASPIILLHKAKEPEITKAVDFDFWA 302

 QY 391 ITDGNKASFLDQVHFQPLPPAVVKLSALAT-ISSAEMKT 430
 DB 303 FKQGRVAA--TELDYVFLPBEVIOIQAQWKTEVKSSDGKT 341

RESULT 8

ID PSTS HAEIN STANDARD; PRT; 334 AA.
 AC P45192;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PSTS OR H11383.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,


```
RESULT 10
PST3 MYCLE
ID PST3 MYCLE STANDARD; PRT; 369 AA.
AC QCBES;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding protein 3 precursor (PBP-3) (Psts-3).
OS PST2 OR PHO2 OR ML2095.
GN Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Required for binding-protein-mediated phosphate
CC transport (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL583924; CAC31050.1; -
CC PR; B87171; B87171.
CC HSP; P06128; IIXH.
CC Leproma; ML2095; -
CC InterPro; IPR006059; SBP_bac_1_1.
CC Pfam; PF01547; SBP_bac_1_1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 369 PHOSPHATE-BINDING PROTEIN 3.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (POTENTIAL).
FT SEQUENCE 369 AA; 37733 MW; 697DECB6A84B9AC CRC64;
Query Match 6.6%; Score 335.5; DB 1; Length 369;
Best local similarity 28.9%; Pred. No. 1.9e-08;
Matches 109; Conservative 50; Mismatches 159; Indels 59; Gaps 13;
QY 75 CGSKPPSGSPETGAGAGTATTASSPV-----TLAETGSLTLYPLFNWGPAPHERYP 128
DB 22 CGS-----DNNAAVGSARTGSSCGVSGGRTPLKASGSTAAQAMTFVNAFERSCP 74
QY 129 NVTITAGTSGAGIAQAAGTGVNIGASDAYLSEGDMAHK-----GLMNIALAISAOV 183
DB 75 GQTLNLTANGSAGVSEFNGNQTFDGGSDPSPLSRKEYAAAEQCGSQAWNLPLVFGPIAI 134
QY 184 NYNLPGVSEHLKNGKVLAAVYOGTITKTDWDDPOIALNPGVNLPGTAVVPLHRSQSGGT 243
DB 135 TYNVNGUSS-LNLDGPTTKIFNGSIASWMDPAIQALNTGVALPASPPIHVFRNDSGGT 193
QY 244 FLFTQYLSKQDPGEWGKSGPGTGTVDPPAVFGALGE--NNGGMVGTGCAETFGCAVYIGI 301
DB 194 DNFORLYDVASNGEWK--GIGKT-----FKGGVGEBAKNGDGTSAAVKSTEGSITNEW 246
QY 302 SFLDQASQORGLGEAQLGNSGNFLLPDAQSIQRAAAGFASKTFAN-----QAI 349
DB 247 SF---ASAKLNTAKIATSAD---PEPIAISVDVSG--KTISGATIIEGENDLVLDTV 296
QY 350 SMIDGPADPGYPIINYEYAIUNNRQKDAATAQLOAFELHWAITDGNKASFLDQVHFQPLP 409
DB 297 SFYKPAQPGSYPIVLTATYIEVCSKYPAQVGRAVKAFLOSTIGGQNG--LGDNQYVPIP 354
QY 410 PAVVKLSDALIATISSA 426
DB 355 -----DSFKSRUSTA 364
DB
RESULT 11
PST2 MYCTU
ID PST2 MYCTU STANDARD; PRT; 370 AA.
AC O05870; P96905;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphate-binding protein 2 precursor (PBP-2) (Psts-2).
GN PST2 OR RV0932C OR MT0959 OR MTCY08D9.07.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Erdmann;
RX MEDLINE=97284492; PubMed=9139906;
RA Lefevre P., Braibant M., de Wit L., Kalai M., Roeper D.,
RA Grotzinger J., Delville J.-P., Peirs P., Ooms J., Huygen K.,
RA Content J.;
RT "Three different putative phosphate transport receptors are encoded by
RT the Mycobacterium tuberculosis genome and are present at the surface
RT of Mycobacterium bovis BCG.";
RL J. Bacteriol. 179:2900-2906(1996).
RN [2]
RP REVISION TO 361.
RA Content J.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann E.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.B., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA DeCher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for binding-protein-mediated phosphate
CC transport.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
```

```
CC CC (Probable).
CC CC -!- INDUCTION: ACCUMULATION OF PROTEIN IS ENHANCED UNDER PHOSPHATE
CC CC STARVATION.
CC CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z48056; CA88137.1; -.
CC CC EMBL; Z95209; CAB08486.1; -.
CC CC EMBL; AE006981; AAK45206.1; -.
CC CC DR PIR; D70584; D70584.
CC CC DR HSSP; P06128; 1LXH.
CC CC DR TIGR; MT0959; -.
CC CC DR Tuberculin; RV0932c; -.
CC CC DR InterPro; IPR006059; SBP_bac_1.
CC CC DR Pfam; PF01547; SBP_bac_1; 1.
CC CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC CC KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
CC CC Complete proteome.
CC CC FT SIGNAL 1 22 POTENTIAL.
CC CC FT CHAIN 23 370 PHOSPHATE-BINDING PROTEIN 2.
CC CC FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
CC CC SQ SEQUENCE 370 AA; 37864 MW; 97F5116CEB9B3B5C CRC64;
CC CC -----
CC CC Query Match 6.5%; Score 330.5; DB 1; Length 370;
CC CC Best Local Similarity 31.1%; Pred. No. 3.1e-08;
CC CC Matches 114; Conservative 44; Mismatches 168; Indels 41; Gaps 13;
CC CC -----
Qy 75 CGSKPPSPGATGAGTAVTTPASSPVTLAETGTTLLPLFLNWPFAFHERYENVVITA 134
Dy 23 CGGTTNSS--SGAG-GTSGSVHCGGKXELHSSSTAGENAVEQPVAYVRSCEGYLDY 79
Qy 135 QGTSGAGIAQAAGTAVNIGASDAYLS-----EGDMAAHK-----GLMNTIALAISAAQVNNYL 187
Dy 80 NANGSGAGVTQFLNNTDFAGSDVPLNPSTQCPDRSAERCGSPAWDLPTVFGPIATYNI 139
Qy 188 PGVSEHLKLVLAAMVQGIKTDQDQIALNPGVNLPGTAVVPLHRSDGSDTLFT 247
Dy 140 KGVST-LMLDPTTAKIPNGITVWNPDPQIALNSGIDLPTTISVIFRSDXGTSNFO 198
Qy 248 QYLSKQDPEGKGFPGFTTVDFFAPVPGALGNGMGVTCAGTGPCVAYIGISFLDQA 307
Dy 199 KYLDGASNGAKG--GASETFNGGVGVGASGNGTSALL---QTGDSITNWSF---A 250
Qy 308 SQRLGEAQLNSSGNFLLPDAQIQAAAGFASKTPANQAI-----SMIDGP 355
Dy 251 VGKQLNQAQIITSAG---PDPVAITTESVG---KTIAGAKIMQGGNDLVDTSFYRPT 303
Qy 356 APDGYPIINYEYAVINRQKDAATAQTLOAFHLWAITDGNKASFLDQVHFOPLPPAV-VK 414
Dy 304 QPGSPVILAYEIVCSKYPDATTGTAVRAFNAQAIQGGQEG--LDQVGSIPLEKSPQAK 361
Qy 415 LSDALIA 421
Dy 362 LAAAVNA 368
CC CC -----
CC CC RESULT 12
CC CC ID PST3_MYCTU STANDARD; PRT; 370 AA.
CC CC AC O86343; Q50794;
CC CC DT 28-FEB-2003 (Rel. 41, Created)
CC CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DE Phosphate-binding protein 3 precursor (PEP-3) (PstS-3) (Antigen Ag88).
CC CC GN PST3 OR PHOS2 OR RV0928 OR MT0955 OR MTCY21C12.22.
CC CC OS Mycobacterium tuberculosis.
```

```
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC STRAIN=Erdbmann;
RX MEDLINE=97000022; PubMed=8843165;
RA Braibant M., Lefevre P., de Wit L., Ooms J., Peirs P., Huygen K.,
RA Wattiez R., Content J.;
RT "Identification of a second Mycobacterium tuberculosis gene cluster
RT encoding proteins of an ABC phosphate transporter.";
RL FEBS Lett. 394:206-212(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RC STRAIN=Erdbmann;
RX MEDLINE=97284492; PubMed=9139906;
RA Lefevre P., Braibant M., de Wit L., Kalai M., Roepers D.,
RA Groetzig J., Delville J.-P., Peirs P., Ooms J., Huygen K.,
RA Content J.;
RT "Three different putative phosphate transport receptors are encoded by
RT the Mycobacterium tuberculosis genome and are present at the surface
RT of Mycobacterium bovis BCG.";
RL J. Bacteriol. 179:2900-2906(1997).
CC -!- FUNCTION: Required for binding-protein-mediated phosphate
CC transport.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -!- INDUCTION: By phosphate starvation.
CC -!- SIMILARITY: BELONGS TO THE PSTS FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z95209; CAB08488.1; -.
CC CC EMBL; Z48057; CA88138.1; -.
CC CC EMBL; AE006981; AAK45202.1; ALT_INIT.
CC CC DR PIR; H70583; H70583.
CC CC DR HSSP; P06128; 1LXH.
CC CC DR TIGR; MT0955; -.
CC CC DR Tuberculin; RV0928; -.
CC CC DR InterPro; IPR006059; SBP_bac_1.
CC CC DR Pfam; PF01547; SBP_bac_1; 1.
```

DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
KW Antigen; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 370 PHOSPHATE-BINDING PROTEIN 3.
FT LIPID 23 370 N-ACYL DIGLYCERIDE (POTENTIAL).
FT VARIANT 112 114 AAA -> PSG (IN STRAIN ERDMAN).
FT VARIANT 116 116 Q -> A (IN STRAIN ERDMAN).
FT VARIANT 118 118 MISSING (IN STRAIN ERDMAN).
FT VARIANT 253 254 OH -> HD (IN STRAIN ERDMAN).
FT SEQUENCE 370 AA; 37953 MW; 75557829A9A118E0 CRC64;
SQ SEQUENCE 370 AA; 37953 MW; 75557829A9A118E0 CRC64;
Query Match 6.2%; Score 314; DB 1; Length 370;
Best Local Similarity 30.8%; Pred. No. 1.7e-07;
Matches 106; Conservative 39; Mismatches 161; Indels 38; Gaps 11;
QY 86 TGAGAGTATTPASSPV-----TLAETGSTALLPLFLNMGPAFHRYENVVITTAQGTGS 139
DB 30 TGGGA---TTQASAKVDCGKTKLKASGSTAQAANMTRFVNVFQACPGQTINTANGS 86
QY 140 GAGIAAAGTGNIGASDAYLSEGDMAHK-----GLAMNIALAISAAQVYNLPGVSEHL 194
DB 87 GAGISEFNGNOTDFGSDVFLSKDEAAAAQRRCGSPAWNLPVVFGLPIAVTYNLSVSS-L 145
QY 195 KLNGKVLAMVQTIKTWDDPQIALNPGVNIPLGTAVPLHRSDSGDTFLTLYLSKQD 254
DB 146 NLDGPTLAKIFNGSITQWNPAIQALNRDFTLFGRIHVFRSDSGTNDNFQRYLQAS 205
QY 255 PEGWGSFPGTGVTPAVPGALGE--NGNGMVTGCAETPGCAVYIGISFLDQASQRL 312
DB 206 NGAWGKAG-----KSGQGVGEGARGNDGTSAAAKNTPGSIYNEWSF---AAQHL 255
QY 313 GEALQNSGNFLPDAQIQAAAGFASKT---PANQIAISMIDG-----PAPDGPPIY 365
DB 256 TMANIVTSAGGD--PVAITIDSVGQTIAGATISGVGNDLVLTDSFYRPRKRGSPYVILA 313
QY 366 EYAIIVNRQKDAATACTLOAFHLWALTIDGNKASFLDQVHFQPLP 409
DB 314 TYEIVCSKYPDSQVGTAVXAFLOSTIGAGQ--SLGDNQYIIP 355
RESULT 13
RAA3 CHLRE
ID RAA3 CHLRE STANDARD; PRT; 1783 AA.
AC OSPECA
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trans-splicing factor Raa3, chloroplast precursor.
GN RAA3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=137C / CC-125;
RX MEDLINE=21161833; PubMed=11285239;
RA Rivier C., Goldschmidt-Clermont M., Roach J.-D.;
RT Identification of an RNA-protein complex involved in chloroplast
RT group II intron trans-splicing in Chlamydomonas reinhardtii.;
RL EMBO J. 20:1765-1773(2001).
CC -!- FUNCTION: Required for trans-splicing of exons 1 and 2 of the
CC chloroplast encoded psaA mRNA (a group II intron). May be required
CC for stability of the chloroplast RNA-protein complex in which it
CC is found.
CC -!- SUBUNIT: Part of a 1700 kDa complex that includes the
CC precursor RNA to exon 1 and the tscA RNA.
CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.
CC -!- DOMAIN: The N-terminal 453 amino acids are dispensable, while the
CC C-terminal 630 amino acids are required for function.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF310675; AAG40000.1; -;
DR EMBL; AF310674; AAG39999.1; -;
KW Chloroplast; Transit peptide; mRNA processing; mRNA splicing.
FT TRANSIT 1 40 CHLOROPLAST (POTENTIAL).
FT CHAIN 41 1783 TRANS-SPICING FACTOR RAA3.
FT DOMAIN 112 311 ALA-RICH.
FT DOMAIN 343 436 SER-RICH.
FT DOMAIN 478 1078 ALA-RICH.
FT DOMAIN 1310 1415 ALA-RICH.
FT DOMAIN 1416 1430 GLN-RICH.
FT DOMAIN 1496 1506 ARG-RICH.
FT DOMAIN 1678 1722 ALA-RICH.
FT DOMAIN 771 778 POLY-ALA.
FT DOMAIN 920 926 POLY-PRO.
FT DOMAIN 927 932 POLY-ALA.
FT DOMAIN 1047 1055 POLY-GLY.
FT DOMAIN 1318 1325 POLY-ALA.
FT DOMAIN 1405 1415 POLY-ALA.
FT DOMAIN 1669 1675 POLY-PRO.
FT SEQUENCE 1783 AA; 180399 MW; 40F6206BA6BBDDB CRC64;
Query Match 4.0%; Score 203.5; DB 1; Length 1783;
Best Local Similarity 20.9%; Pred. No. 0.081;
Matches 226; Conservative 101; Mismatches 425; Indels 327; Gaps 46;
QY 52 SAGKITVRIKLEVSFMRPAQPR-----CGSK-----PPSGFETG-----AGA 90
DB 136 SAAPQTGATPAATAKTTPQRPESDADAGSRAQSQYQFGDPGGGALKGAVDAASDAAP 195
QY 91 GTVATTPASSPVTLAETGSTALLPLFLNMGPAFHRYENVVITTAQGTGSGAGIAQAAAGT 150
DB 196 DVAASAP--PAGISDQLSTPACP-----PEREPQ---AGKPRASGRAPAAAGVGP 241
QY 151 VNTGASDAYLSEGDMAHKGLAMNIALAISAAQVYNLPGVSEHLKNGKVLAAAMYQGTIK 210
DB 242 QDVGGSGCACAPAPDESHMGL-----THRDQGHDERIS-----QTAGE 279
QY 211 TWDDPQIAA--LMPGVNLPCTAVVP-----LHRSQSGDTFLTQVLSKQDEGKG 259
DB 280 AWKAGAVAAPPAATPPPGIAAAPTRLASSALGTHSSDGD-----MRAVP---G 327
QY 260 KSPGFGTVDFFPVPVPGAL-----GENGNGMVTGCAETPGCV--AYIGISFLD- 305
DB 328 RD-----TPSLSAVAGPVTLSGSSSSSSSSSSSSNTSTSTSTNGVTITSNVGVNGASP 382
QY 306 ----QASQRLGEAQ-----LGNSSGNFL-LPDQASIQAAAGFASKTPANQAIISMIDGPA 356
DB 383 QERLMAAFRAVVTQWNTHLGRGRGFAPLPTGGMSTATSAASSSTSSASSSSMNDG-- 440
QY 357 PDGYPIINYEYAVINNRQKDAATAQTALQAFHLWALTIDGNKASFLDQVHFQPLP- 409
DB 441 -----SNAKKTSDAVSLPV-----GQPAEQPHVPTAGGPGSQTG 477
QY 410 -----PAVKLSDALIATISSAEMKTDATLAQEAQNERISG----- 447
DB 478 ASAVAAQAPSSAPTAAMAATMGSAATLPTAAVSVSSAAAEQTQPSGLLAGRPA 537
QY 448 ----DLATQTDQVSTAGSLQGWGAAGTAQAAAVVRFQEAANKQKQELDEISTNIRA 503
DB 538 LLGETIQGRIARLQAREALRAARHARVGAAMQPPV-----QARPVQSGSQVPPV 589
QY 504 GVQYSRADEBQQALSSQMGFTQSVTVVDQBIANRANEVEAPMADPPDVTPIPCETL 563
DB 590 G-----QGFVQSQPRRQEPAAATK---LHVADGLPARVPQPAVSAATDLQDTA 636
QY 564 AAKVAQQLVLSADNMEEYLAAGAKERQRLATSLRNAA-KAYGEVDEEA--ATALONDGE 620

QY 448 DLKTDQID--QVESTAGSLGQWRGAAGTA-----AQAAYVRFQEAANKQKQELDEIST 498
 Db 518 IVEVTTIDGRKXGSLVGAEEVRSIGIGWENLLMEYAVSTMLINWLVDPENDL-QIRC 576
 QY 499 NIR----QAGVQ--YGRADEEQOALSSQM-GFTQSTQTVVQOQILNR-----ANEVEA 546
 Db 577 HIRAQFISGICKRLLKMEGFQVEVDKQIEHFRENE--AIDYEDLLQRESSTKDSIEG 634
 QY 547 PMADPTDYPITPCETAA-----KNAQOQLVLSADN----- 578
 Db 635 EVKD-MTD---PLQITDAIAGRLNTRAHDFVLSALQHLIRNSGDBGRLMYQLVDA 689
 QY 579 MREYLAAGAK-----BRQLATSLRNAKAYGEVDEEATAALDNDGEGTVQAESAGA--- 630
 Db 690 MLSYVAMDRLPLDLRQGLTFVQSLDLRL-HTDAEARRAYDESLEARQIAEAALAERD 748
 QY 631 -----VGDGS-----SAELTDTPRV-ATAGEFNMDL 656
 Db 749 EMKAQVELGADGLVRKLOKQIEBQTGIIELQSRQNMELKAEIADYQRLRAQELQENELET 808
 QY 657 KEA-----ARKLETGDOGASIAHFADGWNFTNL-----TLQ 687
 Db 809 RELYMLRDAQDIAASNAKSNNGEAEETPAHMRGILDRKLLTRLEKQLETKTQFKLE 868
 QY 688 GDV-----KRGRF-DNWEGDAATACEASLDQORQWILHWAKUSAAWAKQAVYQIH 739
 Db 869 GKWVGCHDPSDRILRELREQMDGAGP--REAFEEQARLNLNPNVGSVYRKKT-YIQGME 925
 QY 740 VVARREHPTYED-----IVGLER-----LYAENP-----SARDQILPV 772
 Db 926 DRTATELGQTDDEVVAKARLVDLHRPMDPQATGLLGEIATAKVKIDADDADKDEGRP- 984
 QY 773 YAEYQQRSEKVLTEYN-----NKAALFPVNPKPAPAI-----KIDPPPPPOBQG 817
 Db 985 -TESEQPAEGAAKTEGQGVDDTVAVDKATAAPPPPPPPPAHPLGSLGAAPPPPPPPPP-- 1041
 QY 818 LIPGFLMPSDQSGVTPGTGMPAAAMVPTGSPGGLPADTAQLTSAGREAAALSGDVA 877
 Db 1042 -----PPPPGAGAAPPP--PPPPPPPPPPGLGPPPPPPPPPP----- 1076
 QY 878 VKAASLGGGGGGGVPAPLGSIGAESVRPAGAGDIAGLQGRAGGGAALGG 930
 Db 1077 -----PGFGGPPPPPPPPPGFGGPPPPPPPPPGGAFGVPPPPPPPGTVIGG 1123

RESULT 15

CF10_MYCLE
 ID CF10_MYCLE STANDARD; PRT; 99 AA.
 AC O3084;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 10 kDa culture filtrate antigen cfp10 homolog.
 GN MLO050 OR MLCB628.13C.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 (1)
 RP SEQUENCE FROM N.A.
 RA Bigliemi K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Bigliemi K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy J., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RL "Massive gene decay in the leprosy bacillus."
 CC Nature 409:1007-1011(2001).
 CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMEL; Y14967; CAA75210.1; -;
 DR EMEL; AL583917; CAC29558.1; -;
 DR PIR; T10032; T10032.
 DR Leproma; M0050; -;
 KW Complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFESA8A CRC64;
 BY SIMILARITY.
 Query Match 3.6%; Score 184; DB 1; Length 99;
 Best Local Similarity 39.4%; Pred. No. 0.021;
 Matches 39; Conservative 20; Mismatches 40; Indels 0; Gaps 0;
 QY 426 AEMKTAATLAQAGNFERISGLTKTQIDQVESTAGSLGQWRGAAGTAQAAYVRFQEA 485
 Db 1 AEMITEAAILTQQAQFDQIASGLSQERNFVDSIGSFQNTWEGQAASALGALGRFDEA 60
 QY 486 ANKQKQELDEISTNIRQAGVQYSRADDEEQOALSSQMGF 524
 Db 61 MQDQIROLESIVDKLARSGGNYTKTDDANQLLSSKQNF 99

Search completed: November 21, 2003, 16:04:24

Job time : 13:5909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 48.3698 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-54

Perfect score: 5072

Sequence: 1 MGHHHHHHHVILIGTSPTS.....PANTEAVIGNRRRODSKESK 983

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1552.5	30.6	374	2 Q49590	Q49590 mycobacteri
2	1412	27.8	374	2 Q49589	Q49589 mycobacteri
3	483.5	9.5	363	16 Q8PAF7	Q8PAF7 xanthomonas
4	482.5	9.5	392	16 Q8VND0	Q8VND0 anabaena sp
5	469.5	9.3	363	16 Q8PM56	Q8PM56 xanthomonas
6	460.5	9.1	373	17 Q8YGS6	Q8YGS6 aeropyrum p
7	456	9.0	347	16 Q8YD99	Q8YD99 anabaena sp
8	456	9.0	375	17 Q8U247	Q8U247 pyrococcus
9	452	8.9	336	2 Q49675	Q49675 mycobacteri
10	445	8.8	403	16 Q8RY26	Q8RY26 deincococcus
11	444	8.8	383	16 Q55199	Q55199 synechocyst
12	440	8.7	346	16 Q8Z993	Q8Z993 yersinia pe
13	438	8.6	339	16 Q8PM55	Q8PM55 xanthomonas
14	435	8.6	339	2 Q56833	Q56833 xanthomonas
15	431	8.5	346	2 Q8KY90	Q8KY90 edwardsiell
16	425	8.4	314	2 Q9AML6	Q9AML6 edwardsiell

ALIGNMENTS

RESULT 1

Q49590 PRELIMINARY; PR: 374 AA.
AC Q49590; 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE P8C subunit of ABC transporter.
GN P8TS1B.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.,
RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
RL FEMS Microbiol. Lett. 144:235-240(1996).
DR EMBL; X95538; CAA64784.1; -.
DR HSP; P06128; IIXG.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1_1.
SQ SEQUENCE 374 AA; 38088 MW; A7FD5843860B9D7 CRC64;

Query Match 30.6%; Score 1552.5; DB 2; Length 374;
Best Local Similarity 83.1%; Pred. No. 5.6e-71;
Matches 294; Conservative 25; Mismatches 34; Indels 1; Gaps 1;

Qy 71 AQPCSGKPPSGPFGAGAGTATTPASSPVTLAETGTSTLLPLPLNMGPAFHEYPNV 130
Db 20 AAACGNSGTGAPSQ-GGGPVATTPASSPVTLAETGTSTLLPLPLNMGPAFHEYPNV 78
Qy 131 TTTAQGTGSGAGIAQAAGTWNIGASDYLSEGDMAHKLGMNIALAISAQVYNLPGV 190
Db 79 TTTQGTGSGTGI-SQAAGAVGAIGSDAYLSEGDMAHKLGMNIALAISAQVYNLPGV 138

Q8PAF6 xanthomonas
O32484 enterobacte
Q8XB03 escherichia
Q8XZ75 ralstonia s
Q8CVJ8 escherichia
Q8ZKX5 salmonella
Q8Z2P5 salmonella
Q9URV3 pyrococcus
Q5714 methanococ
Q67815 aquifex aeo
Q8XU83 ralstonia s
Q8XU88 pyrobaculum
Q97ZT8 sulfobus
Q50043 mycobacteri
Q9XG94 burkholderi
Q9AM12 edwardsiell
O53714 mycobacteri
O07267 mycobacteri
Q50099 mycobacteri
P73785 synechocyst
Q93AL4 nodularia s
Q8FMN6 corynebacte
Q9EUS5 streptomyce
Q8G7F1 bifidobacte
Q9KZV9 streptomyce
O33896 peccobacter
Q8NMJ8 corynebacte
Q93AL0 aphanizomen
O07265 mycobacteri

Qy	191	SEHLKNGKVLAA	YGGTITKTWDDPQIAALNPGVNLPGTAVVPLHRS	DGSGDTFLFTQYL	255	
Db	139	TEHIKNGKVL	AGMYNGSYKTNWDPQIAGLPGVNLPGTAVVPLHRS	DGSGDTFLFTQYL	198	
Qy	251	SKQDPGWKSK	PGFGTTVDFFPAVPCALGNGGWTGCAETPCGVAYIGISFLD	QASOR	310	
Db	199	SKQDPGWKSK	PGFGTTVAFTVPVPGALGNGGWTGCAETPCGVAYIGISFLD	QASOR	258	
Qy	311	GLGEAOLGNS	GNFLLPDAQSITQAAAAAGFASKTTPANQAI	SMIDGPADPGYPIINYEYIV	370	
Db	259	GLGEAOLAN	ASDKYLLPDAKSTQAAAAAGFASKTTPANQAI	SLINGPADPGYPIINYEYIV	318	
Qy	371	NNRQKDAAT	QAOTLQAFHLWATIDGNKASFLDQVHFQPLP	PAVVKLSDALIATIS	424	
Db	319	NSQOKDAAT	QAOTLQAFHLWASDGNNSFLDKVHFQPLP	ADVAKLSDAQIAKIS	372	
RESULT 2						
Q49589	ID	Q49589	PRELIMINARY;	PRT;	374 AA.	
AC	Q49589;					
DT	01-NOV-1996	(T-EMBLrel. 01, Created)				
DT	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)				
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)				
DE	Pats	subunit of ABC transporter.				
GN	PST31A.					
OS	Mycobacterium	intracellular.				
OC	Bacteria;	Actinobacteria;	Actinobacteridae;	Actinomycetales;		
OC	Corynebacteri	ineae;	Mycobacteriaceae;	Mycobacterium.		
OX	NCBI_TaxID=1767;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATCC 35761;					
RX	MEDLINE=97055782;	PubMed=8900068;				
RA	Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,					
RA	Moreno C., Ivanyi J.					
RT	"Puplication of genes encoding the immunodominant 38 kda antigen in					
RT	Mycobacterium intracellulare."					
RL	FEMS Microbiol. Lett. 144:235-240 (1996).					
DR	ENBL; X95538; CAA64783.1;					
DR	HSP; P06128; IA54.					
DR	InterPro; IPR006059; SBP bac_1.					
DR	Pfam; PF01547; SBP bac_1.					
SQ	SEQUENCE	374 AA; 37398 MW; 45796D4B9F6F513D CRC64;				
Query Match						
Best Local Similarity		27.8%; Score 1412; DB 2; Length 374;				
Matches		267; Conservative 33; Mismatches 49; Indels 0; Gaps 0;				
Qy	75	CGSKPPSGPETGAGAGTVA	TTTPASSPVTLAETGSTLLYPLFLN	LGPAFHERVNTVTA	134	
Db	24	CGAHATRGAPSTG	TPAGSVATAPATSIETSETGSTLLYPLF	SEWGAAPAKHPNVTTT	83	
Qy	135	QGTGSGAGTAA	AGATVNIIGASDAYLSEGDMAAHKGLM	NIALAISAQVNNY	194	
Db	84	QSGSGAGTSQA	AGAVVIGASDAYLSAGDMAAHKGLM	NIALAISAQVNNY	143	
Qy	195	KLNGKVLAA	MYGGTITKTWDDPQIAALNPGVNLPGTAVVPLHRS	DGSGDTFLFTQYLSKD	254	
Db	144	KLNGKVLAG	MYGGTITKTWDDPQIAGINPGLPATVPVPLHRS	DGSGSTFQFTQYLSKD	203	
Qy	255	PSGWKSKPGFG	TTVDFFPAVPGALGNGGMYTC	CAETPCGVAYIGISFLD	QASORGLGE	314
Db	204	PDGWRSPGFG	TTVSFPVPEALGENDGAMVTGCAANP	CGCVATYTGISLDEADQ	RGLE	263
Qy	315	AOLGNSGNFL	LPDAQSITQAAAAAGFASKTTPANQAI	SMIDGPADPGYPIINYEYIV	VNNRQ	374
Db	264	AKLGWASGYLL	PNAKSTQAAAAAGFASKTTPANQAI	SLINGPATDGTGPIINYEYIV	VNADQ	323
Qy	375	KDAATAQTL	QAFHLWATIDGNKASFLDQVHFQPLP	PAVVKLSDALIATI	423	
Db	324	RDDATAQTL	QAFHLWATIEGNGPSFLDKVHFQPLP	AEVWTLSDAQIAKI	372	

OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Cannavari F., Cardoso J., Chambergo F., Ciapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira C.F., Oliveira V.R.,
Pereira H.A., Rossi A., Seta J.A.D., Silva C., de Souza R.F.,
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463 (2002).
DR EMBL; AE011790; AAM36445.1; -
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1.1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41343 MW; FF28D266FB59C950 CRC64;

Query Match 9.5%; Score 482.5; DB 16; Length 392;
Best Local Similarity 33.7%; Pred. No. 8.5e-17;
Matches 124; Conservative 64; Mismatches 151; Indels 29; Gaps 12;
QY 75 CGSKPPSPGSPGAG-AGTVATTPASSP-----VTLAETGTLPLPLNMGPAFHE 125
DB 29 CGGQGSNTATQDSSGTAKDATASSPAKLDLGGNVSLTGAGSPAPLVSFTDLNK 88
QY 126 RYPNTVITTAQGTGSGAGIAQAAGVTNIGASDAYLSEGDMAAHLNIAIAISAOQVN 184
DB 89 KYENLQINQSVSGAGVEQFTQGTVDGASDVAMKDEIQKVGRLVLLPVTAGGIVLA 148
QY 185 YNLPGVSEHLKNGKVLAMVYGGTITWDDPQIAALNPGVNLPGTAVPLHRSQSGDTF 244
DB 149 YNLPGVTE-LNLPRVYVTDILGLKIKTDWDAPEIKAAFNPNLPSQPTVTVRSQSGSTG 207
QY 245 LFTQYLSKQDPGKWSGPGFTTVDFPVPALGNGNGMVGTCGTPGCAVYIGISFL 304
DB 208 VFTKHLAASVSP-WKSKVGEKSVSNPV---GVGKGNVGTQAIKQTOGAIGVYGY- 262
QY 305 DQASQRLGEAQLGNSSGNFLPLDPAQSTQAAAGFASTPANQAIMSDGPAQD---YP 361
DB 263 --AKQNNISYATLENKAGKFFVYNDESASQTLA--AIQLPENLRAFV---PDPDGDGSGYP 315
QY 362 IINYEVAIVNNRQKDAATQTLQAFHLWAITDGNKASFLDGVHFOPLPAPVVKLSDALIA 421
DB 316 IVSFSIMAYKVPDVRKAKAMEAIEVALTEGQKIS--GELGYIPLQAVVQ-KTATVA 372
QY 422 TISSAEMK 429
DB 373 DQISPEVK 380

RESULT 5
Q8PM56 PRELIMINARY; PRT; 363 AA.
ID Q8PM56;
AC Q8PM56;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ABC transporter phosphate binding protein.
DE PEST OR XAC1577.
GN Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OS Xanthomonas axonopodis (pv. citri).

OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Cannavari F., Cardoso J., Chambergo F., Ciapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira C.F., Oliveira V.R.,
Pereira H.A., Rossi A., Seta J.A.D., Silva C., de Souza R.F.,
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463 (2002).
DR EMBL; AE011790; AAM36445.1; -
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1.1.
KW Complete proteome.
SQ SEQUENCE 363 AA; 37641 MW; 1A39C0289BF7DBDF CRC64;

Query Match 9.3%; Score 469.5; DB 16; Length 363;
Best Local Similarity 32.2%; Pred. No. 3.5e-16;
Matches 119; Conservative 69; Mismatches 157; Indels 25; Gaps 11;
QY 52 SAGKITYIKIEVSEKWPAPQRCGSKPPSPGSPGAGTVATTPASSPVTLAETGTL 111
DB 9 TALSUTIALAALAAACSPGKDAQSADAQK---GAPAAAGAAAD-----SKAEISGAGASF 59
QY 112 LYPLENLMGPAFHRYPNVITTAQGTGSGAGIAQAAGVTNIGASDAYLSEGDMAAHLN 171
DB 60 IYPLVSKVSADYNAATGN-KVNYQSIGSGGGIAQIKAGTVDFGTDKPLDSAEL-QQAGL 117
QY 172 MNIALAISAOQVYNLPGVSE-HLKNKGVLAMVYGGTITWDDPQIAALNPGVNLPGTA 230
DB 118 GQFPAIGAIGVVPVNLNLEGIAFPKRLTGALLGDLFGKVTVMNDAAIVAANPGVTLPATK 177
QY 231 VYPLHRSQSGDTLFTQYLSKQDPGKWSGPGFTTVDFPVPALGNGNGMVGTCGA 290
DB 178 INLVHRSQSGDTTFNFSNLYSKVSP-WKSKVGEKTSQVW---FGVGGKNGEVASVYQ 233
QY 291 ETPGCVAYIGISFLDQASQRLGEAQLGNSSGNFLPLDPAQSTQAAAG--FASKTPANQA 348
DB 234 QIKGSGIVYELAY---ALQNKMPYTSLQNAAGQWIEPNAESFAAASADANARDNVLV 290
QY 349 ISMIDGPAQDPGPIINYEVAIVNNRQKDAATQTLQAFHLWAITDGNKASFLDGVHFOPL 408
DB 291 IT--NAPGEKAWPITATNFMMLHKQPKDARSKATLDFKFWALENGQAQA--SELHYVPL 346
QY 409 PFAVVKLSDA 418
DB 347 PPELVKQTEA 356

RESULT 6
Q9YGS6 PRELIMINARY; PRT; 373 AA.
ID Q9YGS6;
AC Q9YGS6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 373AA long hypothetical phosphate-binding periplasmic protein.
GN APE0045.
OS Aeropyrum pernix.


```
QY 139 SGAGTAAAGTAVNIGASD-----AVLSEGDMAAHKGLMNLIALISAQOQVNNLPG 189
Db 78 SCHQBAFLKGLTIDGRTPPVKESTWKLETGDQP-----LQPIVGVAVVWHNPN 132
QY 190 VSEHLKNGKVLAAWYOGTITKTWDDPQIAALNPGVNLPGTAVVPLHRSDDSGDTFLFTQY 249
Db 133 VDE-LKLDGETLAKIFMGQIEYDDPAIKSLNPNVNLPHKEIIIVHRSDDSGTITAITTY 191
QY 250 LSKODPEGWGSFGFTTVDPPA--VFGALGNGNGWVTGCASTPCGVAYIGISFLDQA 307
Db 192 LSLVSEB-WATKVAGKTVDPVQKVGRIQAKNPGVQQLKNTPISIAYTELSP---A 247
QY 308 SQRGELGALQNSNGNFFLLPDAQSIQAAAAGFASKTP-----ANQAI SMIDGPADPGYP 361
Db 248 IEENLKI VALKXKAGNFVKPTEETIKAAVAAVKAFIPDPAEGYKEDIRQLLNAPGNSYP 307
QY 362 IINVEYATV-----NNRQKDAATQTLQAFHWAITDGNKASFLDQVHFQPLPRAVVKL 415
Db 308 IVATFHLVWQNGKGHYSKKAQAIKDFLRWLTEGQKPENI-APGVGLPPEVAEI 364

RESULT 9
Q49675
ID Q49675 PRELIMINARY; PRT; 336 AA.
AC Q49675;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PabB.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96084954; PubMed=7489918;
RA Doukhan L., Predich M., Nair G., Dussurget O., Mandic-Mulec I.,
RA Cole S.T., Smith D.R., Smith I.;
RT "Genomic organization of the mycobacterial sigma gene cluster.";
RL Gene 165:67-70(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; J00012; AAA85925.1; -.
SQ SEQUENCE 336 AA; 35916 MW; 0C7CABDB3CBD99F CRC64;

Query Match 8.9%; Score 452; DB 2; Length 336;
Best Local Similarity 40.1%; Pred. No. 2.4e-15;
Matches 122; Conservative 22; Mismatches 74; Indels 86; Gaps 11;

QY 93 VATTPASSPVTLAETGSLTLLYPLFNWGP-AFHERYPNVTITTAQGTGSGAGIAQAAGTV 151
Db 39 VTTATATSKVTLSKTGSLTLLNPLNLSYLLRSKYPNVTITQGT----GIAQAAGTV 94
QY 152 NIGASDAYLSEGDMA-AHK-----GLMNLIALISAQOQVNNLPGVSEHLKNGKVLAAWYQ 206
Db 95 DIGVSDAYFTVGRHNSNAHKIDSHRYSHLSANQLGSVRHQAPQTEWQAG-----AAATQ 150
QY 207 GTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDDSGDTF---LFTOYLSKQDPEGWG---- 259
Db 151 GTIKTWNDPQITALNPGVNLPGTAVVPLHLSDDGSDNTLLVHLTHYLAQPDGNGISRL 210
QY 260 -----KSPGF-----GTTVDPAVPGALGNGNGWVTGCAET 292
Db 211 ASARPTSPSFAHWAKQQRWHSYQRLRLQAASTSTPTRRGGARANRAMPVFNCCST 270
QY 293 PGCVAYIGISFLDQASORGLGEAQLGNSNGNFFLLPDAQSIQAAAAGFASKTP-ANQAI SM 351
Db 271 PKV-----FRLQPAS-----PRTPRRSKVTHL 293
QY 352 IDGP 355
|||||
```

```
Db 294 IDGP 297
RESULT 10
Q9RYZ6 PRELIMINARY; PRT; 403 AA.
AC Q9RYZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Phosphate ABC transporter, periplasmic phosphate-binding protein.
GN DRA0157.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001862; AAF12207.1; -.
DR HSP; P06128; IIXG.
DR TIGR; DRA0157; -.
DR InterPro; IPR006059; SBP_bac.1.
DR Pfam; PF01547; SBP_bac.1; 1.
DR Complete proteome.
KW SEQUENCE 403 AA; 42060 MW; 38DC61C6E7D290D0 CRC64;

Query Match 8.8%; Score 445; DB 16; Length 403;
Best Local Similarity 31.5%; Pred. No. 6.9e-15;
Matches 119; Conservative 60; Mismatches 157; Indels 40; Gaps 11;

QY 73 PRGSKPPSGS-----PETGAGAGT---VATTPASSPVTLAETGSLTLLYP 114
Db 40 PLCGTSPVFGSLDAPLTKEPHTRRMKKTLGLSALVMISTAAQAIGTA--GASFPYP 97
QY 115 LFNLMGPAPHERYPNVTITTAQGTGSGAGIAQAAGTVNIGASDAYLSEGDMAAHKGLMN 173
Db 98 LYS-----KMFSEYKASNNVYQSVGSGSGQKILERTVDFAGSDNPNMTDAQLGSAPGTL 153
QY 174 IALAI SAQOQVNNLPGVSEHLKNGKVLAAWYOGTITKTWDDPQIAALNPGVNLPGTAVVP 233
Db 154 VPTAIGAVPAYNLPVGTKEPLNFDGPTLANIYLGKIKTWGDDPALAKLNGVTIPLPIV 213
QY 234 LHRSDSGDTFLFTOYLSKODPEGWGSFGFTTVDPAVPGALGNGNGWVTGCAETP 293
Db 214 ARSSDGGSTTFVFDYLSKVSGE-WKSKVGAGNSLOWPVGTGAKGNDGVAGVVKG---TP 269
QY 294 GCVAYIGISFLDQASORGLGEAQLGNSNGNFFLLPDAQSIQAAAAGFASKTPANQAI SMID 353
Db 270 GALTGYVELVY---AKQNKJUSFGAVKRRAGKFIADNGPASNALGVV--IPADTRVSLTN 334
QY 354 GPAPDGPYIINVEYATVNNRQK-----DAATAQTLQAFHWAITDGNKASFLDQVHFQPLP 409
Db 325 SANAGAYPIASFTYLYFYKQKYGNRTAQAKALKULLTVTVVTSQQQ--YNEGLDYAKLP 382
QY 410 PAVVKUSDALIAIIS 424
Db 383 SNVAAKAKTIINSMN 397

RESULT 11
Q55199
```

Q8Z9S9	PRELIMINARY;	PRT;	346 AA.
ID	Q8Z9S9		
AC	Q8Z9S9		
DT	01-MAR-2002 (T-EMBLrel. 20, Created)		
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)		
DE	Putative phosphate-binding periplasmic protein (Periplasmic		
DE	phosphate-binding protein).		
GN	PTS OR YPO4117 OR Y4131.		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Yersinia.		
NCBI_TaxID=632;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CO-92 / Biovar Orientalis;		
RX	MEDLINE=21470413; PubMed=11586360;		
RA	Parkhill J.J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,		
RA	Prentice M.B., Sebathia M., James K.R.D., Churcher C., Mungall K.L.,		
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,		
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,		
RA	Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,		
RA	Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,		
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;		
RT	"Genome sequence of Yersinia pestis, the causative agent of plague.";		
RL	Nature 413:523-527(2001).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=KIM5 / Biovar Mediaevalis;		
RX	MEDLINE=22137863; PubMed=12142430;		
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,		
RA	Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,		
RA	Fetherston J.D., Lidlner L.E., Brubaker R.K., Plano G.V.,		
RA	Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,		
RA	Perry R.D.;		
RT	"Genome sequence of Yersinia pestis KIM.";		
RL	J. Bacteriol. 184:4601-4611(2002).		
DR	EMBL; AJ414160; CAC93566.1; -		
DR	EMBL; AB014014; AAM87673.1; -		
DR	InterPro: IPR006059; SBP_bac_1.		
DR	Pfam: PF01547; SBP_bac_1; 1_1.		
KN	Hypothetical protein; Complete proteome.		
SQL	SEQUENCE 346 AA; 36710 MW; 0C9E4FCF55D3C181 CRC64;		
Query Match	8.7%; Score 440; DB 16; Length 346;		
Best Local Similarity	33.4%; Pred. No. 1e-14;		
Matches 110;	Conservative 60; Mismatches 141; Indels 18; Gaps 10		
QY	90 AGTVATTAS--SPVTLAETGSTLLYPLFNWLPAPHERYPNVTITTAQGTSGAGIAQAA 147		
Db	13 AATLSMTAVSAFAASLTGAGTAPPAVYAKWADSVQKGTGN-KINYQIGSGGGVKQII 71		
QY	148 AGTVNIGASDAYLSEGDMAAHKGLMNLIALAISAQVNYNLPGV-SEHLKLNKVLAAMTQ 206		
Db	72 ANTVDFGASDAPLTDEKLT-EGLFQFTVIGGVLAIVNIPGKSGELTLDGKTLGDYIL 130		
QY	207 GTIKTWDDPQIAALNPGVNLPGTAVPLHRSDCSGDTFLFTQVLSKODPEGWKSGEFGT 266		
Db	131 GTVKKNDPQPAIVKLNPGVKVLPQDNIAVVRADSGSTSFVFTSLAKNAE-WREKVGAGS 189		
QY	267 TVDFPAPVFGALGNGNGWVTGCATPGCVAYIGISFLDAQSGRGLGEAQLGNSSGNFLL 326		
Db	190 TVNWPT---GGGKGNDCIAAFVORLPGSTGYEYAV--AKQNNLAYTKLISADGKPKVS 243		
QY	327 PDAQSIQAAAAGFA-SKTPANQAISMTIDGAPPGYPIIAVVEAVINNRKDKDAAQTQLCA 385		
Db	244 PTEHSFSSAAKGVDSKSFQ---DLINQKDDVWFTITSTFILVHKERKNAANGTEVLK 300		
QY	386 FLHWAITDGNKASFLDQVHFQPLPPAYVK 414		
Db	301 FFDWGYTHGARQA--NELDYATLPAEYVE 327		

RESULT 12

RESULT 13

Q8PM55 PRELIMINARY; PRT; 339 AA.
 ID Q8PM55
 AC Q8PM55
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Phosphate binding protein.
 GN PHOX OR XAC1578.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2202145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.M.A., Reinach F.C., Farah C.S., Furian L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fotmighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AE011790; AAM36446.1; --
 DR InterPro; IPR005673; Peri-phosph.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR TIGRFAMs; TIGR00975; 3a0107s03; 1.
 ZW Complete proteome.
 SQ SEQUENCE 339 AA; 35623 MW; 90E16808A4A710B1 CRC64;

Query Match 8.6%; Score 438; DB 16; Length 339;
 Best Local Similarity 32.1%; Pred. No. 1.2e-14;
 Matches 114; Conservative 60; Mismatches 153; Indels 28; Gaps 9;

QY 65 SFKRPAPRCGSKPPSPETGAGTAVTTPASSPVTLAETGTLPLFLNLWGPAPH 124
 DB 4 SFKRLA-----VGVLAAALCAQAADVTGA--GASFTYPMVMSKWS-ADY 46
 QY 125 ERYPNVTITTAQGTGAGIAQAAGTAVNIGASDAVLSGDMAAHKLMLNIALAISAOQVN 184
 DB 47 NNAATKQVNYQSIGSGGIAQIKAAASVDFGSSDAPLKPEELAA-AGLAQFSPVIGVVPV 105
 QY 185 YNLPGVSE-HLKLNGKVAAMVQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGSDT 243
 DB 106 VNVPGIAAGTLKLDGTLGDFLGVSTWNPDAALNPGVNLPGTAVVPLHRSDSGSDT 165
 QY 244 FLFTQYLSKQDPGEGWKSFGFTTVDFFAVPGALGNGMGVTCGAETPGCVAYIGISF 303
 DB 166 FNFTYLSKVNPD-WKGVGEGTAVQWPT---GIGKGNEGVAAYVKQIKGGIGYVELSY 221
 QY 304 LDQASQRLGEAQLGNSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAFGVPII 363
 DB 222 ---ALQNKVAYTAMKNAAGKVFQPSDETFAAANSADWGSKDFYLVMTNAGDNWAPIT 278
 QY 364 NYEYAIYNNRQKDAATAQTQLAFHLWAITDGNKASFLDQVHFQPLPPAVVKSLSDA 418
 DB 279 ATNFILVQKKPKNPAGLKNLTLEFFRWYTKGDAQA--KOLDYVPLDPLVTQIEA 331

RESULT 14

Q56833 PRELIMINARY; PRT; 339 AA.
 ID Q56833
 AC Q56833
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Phosphate binding protein.
 GN PHOX.
 OS Xanthomonas oryzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=347;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96063021; PubMed=7579584;
 RA Hopkins C.M., White F.F., Heaton L.A., Guikema J.A., Leach J.E.;
 RT "A homolog of an Escherichia coli phosphate-binding protein gene from
 RT Xanthomonas oryzae pv.oryzae";
 RL DNA Seg. 5:299-305(1995).
 DR EMBL; X81478; CAA57231.1; --
 DR HSSP; P06128; 2ABH.
 DR InterPro; IPR005673; Peri-phosph.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR TIGRFAMs; TIGR00975; 3a0107s03; 1.
 SQ SEQUENCE 339 AA; 35632 MW; B15241EF192ED72E CRC64;
 Query Match 8.6%; Score 435; DB 2; Length 339;
 Best Local Similarity 32.0%; Pred. No. 1.7e-14;
 Matches 115; Conservative 60; Mismatches 156; Indels 28; Gaps 9;
 QY 65 SFKRPAPRCGSKPPSPETGAGTAVTTPASSPVTLAETGTLPLFLNLWGPAPH 124
 DB 4 SFKRLA-----VGVLAAALCAQAADVTGA--GRSFTYPMVMSKWS-ADY 46
 QY 125 ERYPNVTITTAQGTGAGIAQAAGTAVNIGASDAVLSGDMAAHKLMLNIALAISAOQVN 184
 DB 47 NNAATKQVNYQSIGSGGIAQIKAAASVDFGSSDAPLKPEELAA-AGLAQFSPVIGVVPV 105
 QY 185 YNLPGVSE-HLKLNGKVAAMVQGTIKTWDDPQIAALNPGVNLPGTAVVPLHRSDSGSDT 243
 DB 106 VNVPGIAAGTLKLDGTLGDFLGVSTWNPDAALNPGVNLPGTAVVPLHRSDSGSDT 165
 QY 244 FLFTQYLSKQDPGEGWKSFGFTTVDFFAVPGALGNGMGVTCGAETPGCVAYIGISF 303
 DB 166 FNFTYLSKVNPD-WKGVGEGTAVQWPT---GIGKGNEGVAAYVKQIKGGIGYVELSY 221
 QY 304 LDQASQRLGEAQLGNSGNFLPDAQSIQAAAGFASKTPANQAISMIDGPAFGVPII 363
 DB 222 ---ALQNKVAYTAMKNAAGKVFQPSDETFAAANSADWGSKDFYLVMTNAGDNWAPIT 278
 QY 364 NYEYAIYNNRQKDAATAQTQLAFHLWAITDGNKASFLDQVHFQPLPPAVVKSLSDA 422
 DB 279 ATNFILVQKKPKNPAGLKNLTLEFFRWYTKGDAQA--KOLDYVPLDPLVTQIEAYWAT 335
 RESULT 15
 Q8KY90 PRELIMINARY; PRT; 346 AA.
 ID Q8KY90
 AC Q8KY90
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Phosphate-binding protein.
 GN PSTs.
 OS Edwardsiella tarda.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Edwardsiella.
 OX NCBI_TaxID=636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tan Y.P., Leung K.Y.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```

DR  ENBL; AF248495; AAM45393.1; -.
SQ  SEQUENCE 346 AA; 36264 MW; 9DA3670A6684B821 CRC64;

Query Match      8.5%; Score 431; DB 2; Length 346;
Best Local Similarity 35.7%; Pred. No. 2.9e-14;
Matches 121; Conservative 50; Mismatches 136; Indels 32; Gaps 14;

QY  87 GAGACTV--ATTPASSPTVLAETGTLVFLNLWGPAPHERYPNVTTAQTGSGAGIA 144
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  145 QAAQTGTVNIGASDAYLSEGDMAAHKLMNLTALAISAQVYNLPGV-SEHLKLNKVLAA 203
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  69 QIANTVDFGASDAPLSDEKLA--DGLFQPPTVIGVMAVNLPGVKSGELTLDSETLGD 127
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  204 MYCGTIKTMDDPQIAALNPGVNLPGTAVPLHRSDCSGDTPLFTQYLSKQDPGWGKSPG 263
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  128 IYLGKIKWNDAALAKNLPGVKLPDQNIAYVRADSGSTSFVFTSYLSKVNQP-WKADVG 186
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  264 FQTTVDFPAPCALGKNGNGMTWCGAETPCGVAYIGISFLDOASORGIEAQLGNSSGN 323
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  187 AGSTVNWFT---GLGKGKNDGVAAFVORLPDSIGVEYAY---AKONKLYATKVLUSDAGQ 240
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  324 FLLPDAQSIQAAAAG-----FASKTPANQAI SMIDGPADGPIYINTEYAI VNNRQ-KD 376
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  241 AVSPTEVSFSNAAGADWSRSFAQDL-TNQA-----GDNVWPITSTFTFLVHKATKP 292
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  377 ATAQTQQAFLHWAITDGNK-ASFLDQVHFQPLPPAVWK 414
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  293 KQTAEVLK-FFDWAYNDGAKQAALD---YATLPASVWE 327
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:50:31 ; Search time 32.6333 Seconds
(without alignments)
2621.664 Million cell updates/sec

Title: US-09-688-672a-64

Perfect score: 2737

Sequence: 1 MHHHHHTAASDNFQLSQGG.....RAPVZADAGGQKVLVRNV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

-Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2737	100.0	543	22 AAU01905	M. tuberculosis an
2	1963	71.7	392	20 AAW73765	M. tuberculosis an
3	1963	71.7	392	20 AAW73655	M. tuberculosis an
4	1963	71.7	392	22 AAU08226	Polypeptide encode
5	1963	71.7	392	22 AAU01882	M. tuberculosis an
6	1963	71.7	392	23 AAU05959	M. tuberculosis an
7	1963	71.7	392	23 AAE29714	Mycobacterium sp.
8	1963	71.7	392	23 AA17578	Mycobacterium spec
9	1941	70.9	408	22 AAU01886	M. tuberculosis an

10	1941	70.9	408	22 AAU01887	M. tuberculosis an
11	1718	62.8	788	22 AAU01903	M. tuberculosis an
12	1263	46.1	394	23 ABU05402	M. tuberculosis an
13	1168	42.7	242	22 AAU01883	M. tuberculosis an
14	1060	38.7	744	22 AAU01902	M. tuberculosis an
15	1060	38.7	815	22 AAU01904	M. tuberculosis an
16	1055	38.5	219	22 AAU01884	M. tuberculosis an
17	809.5	29.6	729	20 AAY32059	Mycobacterium tube
18	809	29.6	726	23 AAU74588	Antigenic fusion p
19	809	29.6	729	22 AAO22142	Ral2-H9-32A fusion
20	809	29.6	729	23 AAE29708	Mycobacterium sp.
21	809	29.6	729	23 AAE29709	Mycobacterium sp.
22	809	29.6	729	23 AAE17572	Mycobacterium spec
23	809	29.6	729	23 AAE17573	Mycobacterium spec
24	809	29.6	930	23 AAE29731	Mycobacterium sp.
25	765	28.0	577	22 AAO22140	Ral2-WT1 fusion pr
26	749.5	27.4	220	22 AAO22141	Ral2-mammaglobin f
27	739	27.0	231	20 AAY32071	Mycobacterium tube
28	739	27.0	234	23 AAU74600	Antigenic fusion p
29	737.5	26.9	585	22 AAG83277	Chlamydia trachoma
30	737.5	26.9	585	23 ABB94248	Chlamydia trachoma
31	728	26.6	583	22 AAG83281	Chlamydia trachoma
32	728	26.6	583	23 ABB94252	Chlamydia trachoma
33	727	26.6	230	22 AAO22139	Ral2-DPPD fusion p
34	722.5	26.4	518	22 AAG83276	Chlamydia trachoma
35	722.5	26.4	518	23 ABB94247	Chlamydia trachoma
36	718.5	26.3	529	23 AAU74387	Breast tumour-spec
37	717.5	26.2	525	21 AAB13645	C. pneumoniae sero
38	717.5	26.2	525	22 AAG83213	Protein encoded by
39	717.5	26.2	525	23 ABB94184	Chlamydia protein
40	709.5	25.9	715	22 AAG83273	Chlamydia trachoma
41	709.5	25.9	715	23 ABB94244	Chlamydia trachoma
42	708	25.9	715	22 AAG83275	Chlamydia trachoma
43	708	25.9	715	23 ABB94245	Chlamydia trachoma
44	706.5	25.8	384	23 AAU85596	Ral2-L985P fusion
45	706.5	25.8	384	24 ABU69568	Human Ral2-L985P f

ALIGNMENTS

RESULT 1	AAU01905	AAU01905 standard; Protein; 543 AA.
ID	AAU01905	
XX	XX	
AC	AAU01905;	
XX	XX	
DT	29-AUG-2001 (first entry)	
XX	XX	
DE	M. tuberculosis antigen HTCC#1 fusion protein #4.	
XX	XX	
KW	TbRA12-HTCC#1; antigen; vaccine; tuberculosis;	
KW	AIDS; acquired immunodeficiency disease; His Tag.	
XX	XX	
OS	Mycobacterium tuberculosis.	
OS	Synthetic.	
XX	XX	
FH	Key	Location/Qualifiers
FT	Binding-site	3..8
FT		/label= Histidine tag
FT		/note= "Nickel chelating region used to aid
FT		purification of the protein"
FT	Misc-difference 541	/label= OTHER
FT		/note= "In frame STOP codon"
XX	XX	
PN	WO2000124820-A1.	
XX	XX	
PD	12-APR-2001.	
XX	XX	
PF	10-OCT-2000; 2000WO-US28095.	
XX	XX	
PR	07-OCT-1999; 99US-0158338.	

PR 07-OCT-1999; 99US-0158425.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 PI WPI; 2001-290576/30.
 DR
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 XX
 XX Claim 17; Fig 8; 168pp; English.
 XX
 CC The sequence represents Mycobacterium tuberculosis fusion protein,
 CC TbrA12-HTCC#1 and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous
 CC antigens, as a fusion protein, and vectors expressing the fusion
 CC proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 2737; DB 22; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.2e-219; Indels 0; Gaps 0;
 Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHTAASDNFQSGGQGFALPICOAMAIAGQIRSGGSPVHIGPTAFILGLGVD 60
 Db 2 MHHHHTAASDNFQSGGQGFALPICOAMAIAGQIRSGGSPVHIGPTAFILGLGVD 61
 QY 61 NNGCARVORVVGSPAPASLGISTGDVITAVDGPINSATAMALNGHHHGGDVISTWQ 120
 Db 62 NNGCARVORVVGSPAPASLGISTGDVITAVDGPINSATAMALNGHHHGGDVISTWQ 121
 QY 121 TKSGGTRGNVTLAGPPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGGLY 180
 Db 122 TKSGGTRGNVTLAGPPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGGLY 181
 QY 181 SLEVFKEALEELAAFPDGLWLSAADKYAGKRNHNVNFFQELADLDRLQISLIHQAN 240
 Db 182 SLEVFKEALEELAAFPDGLWLSAADKYAGKRNHNVNFFQELADLDRLQISLIHQAN 241
 QY 241 AVQTRDILEGAKGLEFVRPVAVDLTYIPVVGHALSAFOAPFCAGAMAVVGGALAYLV 300
 Db 242 AVQTRDILEGAKGLEFVRPVAVDLTYIPVVGHALSAFOAPFCAGAMAVVGGALAYLV 301
 QY 301 VKTLINATOLLKLIKLAELVAATAIISDVADIIGKILGEVWFEITNALNGKELWDX 360
 Db 302 VKTLINATOLLKLIKLAELVAATAIISDVADIIGKILGEVWFEITNALNGKELWDX 361
 QY 361 LTGWVTGLFSRGSWNSLEFFAGVPLGTGATSLGQVTLGFGAAGLSASSGLAHADSLS 420
 Db 362 LTGWVTGLFSRGSWNSLEFFAGVPLGTGATSLGQVTLGFGAAGLSASSGLAHADSLS 421
 QY 421 ASLPALAGIGGSGFGGLPFLSAQVHAASSTROALPRADGPVGAAGAEVGGOSQLVSAQGS 480
 Db 422 ASLPALAGIGGSGFGGLPFLSAQVHAASSTROALPRADGPVGAAGAEVGGOSQLVSAQGS 481
 QY 481 QMGGPGVGMGNHPSGASKGTTTKYSEGAAGTDAERAPVEDADAGGQKVLVRNV 539
 Db 482 QMGGPGVGMGNHPSGASKGTTTKYSEGAAGTDAERAPVEDADAGGQKVLVRNV 540
 RESULT 2
 ID AAW73765
 XX AAW73765 standard; Protein; 392 AA.
 AC AAW73765;
 XX

DT 24-MAR-1999 (first entry)
 XX M. tuberculosis antigen clone hTcc#1 protein sequence.
 DE
 XX Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 KW infection.
 XX Mycobacterium tuberculosis.
 OS
 XX WO9853075-A2.
 FN
 XX 26-NOV-1998.
 PD
 XX 20-MAY-1998; 98WO-US10407.
 PF
 XX 05-MAY-1998; 98US-0073010.
 PR
 XX 20-MAY-1997; 97US-0859381.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;
 PI WPI; 1999-045314/04.
 XX N-PSDB; AAX01177.
 DR
 XX Polypeptide comprising immunogenic Mycobacterium tuberculosis
 PT antigen - useful for immunisation against M. tuberculosis infection
 PT to treat or prevent tuberculosis, and in diagnosis of tuberculosis
 XX
 PS Claim 2; Page 87-88; 100pp; English.
 XX
 CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans),
 CC e.g. they can be included with an acceptable carrier in pharmaceutical;
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques.
 XX
 SQ Sequence 392 AA;
 Query Match 71.7%; Score 1963; DB 20; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGGLYSSLEYFEKALEELAAFPDGLWLSAA 207
 Db 1 MSRAFIIDPTISAIDGLYDLIGIPNQGGLYSSLEYFEKALEELAAFPDGLWLSAA 60
 QY 208 DKYAGKRNHNVNFFQELADLDRLQISLIHQANAVQTRDILEGAKGLEFVRPVAVDLT 267
 Db 61 DKYAGKRNHNVNFFQELADLDRLQISLIHQANAVQTRDILEGAKGLEFVRPVAVDLT 120
 QY 268 YIPVVGHALSAFOAPFCAGAMAVVGGALAYLVVKTINATOLLKLIKLAELVAATAIAD 327
 Db 121 YIPVVGHALSAFOAPFCAGAMAVVGGALAYLVVKTINATOLLKLIKLAELVAATAIAD 180
 QY 328 IISDVADIIGKILGEVWFEITNALNGKELMDKLTGWVTGLFSRGSWNSLEFFAGVPLGT 387
 Db 181 IISDVADIIGKILGEVWFEITNALNGKELMDKLTGWVTGLFSRGSWNSLEFFAGVPLGT 240

QY 388 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 300
QY 448 STRQALRPADGPVGAAAEQVGGQSOLVSAQSGQMGPPVGMGMHPSSGASKGTTTKY 507
DB 301 STRQALRPADGPVGAAAEQVGGQSOLVSAQSGQMGPPVGMGMHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDEAPVEADAGGQKVLVRNV 539
DB 361 SEGAAAGTDEAPVEADAGGQKVLVRNV 392

RESULT 3

AAW73655
ID AAW73655 standard; Protein; 392 AA.

XX AC AAW73655;
XX DT 24-MAR-1999 (first entry)

XX M. tuberculosis antigen clone htcc#1 protein sequence.
XX Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.

XX Mycobacterium tuberculosis.
XX WO9853076-A2.
XX 26-NOV-1998.

XX 20-MAY-1998; 98WO-US10514.
XX 05-MAY-1998; 98US-0073009.
XX 20-MAY-1997; 97US-0858998.

XX (CORI-) CORIXA CORP.
XX Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;
XX WPI; 1999-045315/04.
XX N-PSDB; AAX01143.

XX New isolated Mycobacterium tuberculosis antigens - used to develop
XX products for the prevention, treatment and diagnosis of tuberculosis
XX infection
XX Claim 2; Page 88-89; 104pp; English.

XX This sequence represents an immunogenic portion of a Mycobacterium
XX tuberculosis antigen of the invention. The polypeptides are useful for
XX immunotherapy to treat or prevent tuberculosis (especially in humans),
XX e.g. they can be included with an acceptable carrier in pharmaceutical
XX compositions or included in vaccines, and administered to induce
XX protective immunity in a patient against M. tuberculosis. Tuberculosis is
XX a chronic, infectious disease generally caused by M. tuberculosis
XX infection, and if left untreated typically results in serious
XX complications and death. Fusion proteins containing the antigen, or DNA
XX molecules can similarly be included with an acceptable carrier in
XX pharmaceutical compositions or in vaccines and administered as above. The
XX polypeptides are also useful for diagnosis of tuberculosis, by contacting
XX dermal cells with at least one polypeptide and detecting an immune
XX response (especially induration) on the patient's skin. Inhibiting the
XX spread of tuberculosis requires vaccination and accurate diagnosis, since
XX antibiotic therapy may not be effective due to the existence of an
XX asymptomatic but contagious stage and to patient non-compliance. The
XX polypeptides overcome concerns of safety and efficacy of current
XX vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
XX lack of sensitivity and specificity of existing diagnostic techniques.

XX Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 20; Length 392;
Best Local Similarity 99.7%; Pred. No. 3e-155;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 148 MSRAPIIDPTISADGLYDLIGIGIPNQGGLYSSLEYFEKALEELAAAPFGDGLGSA 207
DB 1 MSRAPIIDPTISADGLYDLIGIGIPNQGGLYSSLEYFEKALEELAAAPFGDGLGSA 60
QY 208 DKYAGKRNHNFFFOELADLDRLQLISLIHDOANAVQTTTRDILEGAKKGLFVRPVAVDLT 267
DB 61 DKYAGKRNHNFFFOELADLDRLQLISLIHDOANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 327
DB 121 YIPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 180
QY 328 IISDVADIIKGLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAGVPGLT 387
DB 181 IISDVADIIKGLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAGVPGLT 240
QY 388 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA 300
QY 448 STRQALRPADGPVGAAAEQVGGQSOLVSAQSGQMGPPVGMGMHPSSGASKGTTTKY 507
DB 301 STRQALRPADGPVGAAAEQVGGQSOLVSAQSGQMGPPVGMGMHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDEAPVEADAGGQKVLVRNV 539
DB 361 SEGAAAGTDEAPVEADAGGQKVLVRNV 392

RESULT 4

AAU08226
ID AAU08226 standard; Protein; 392 AA.

XX AC AAU08226;
XX DT 17-DEC-2001 (first entry)

XX Polypeptide encoded by Mycobacterium tuberculosis clone htcc#1.
XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX immunostimulant; clone htcc.

XX Mycobacterium tuberculosis.
XX WO2000162893-A2.
XX 30-AUG-2001.

XX 26-FEB-2001; 2001WO-US05992.
XX 25-FEB-2000; 2000US-0185037.
XX 08-AUG-2000; 2000US-0223828.

XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Skeiky Y, Ovendaie P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX N-PSDB; AAS12487.

XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX response to and inhibiting development of a Mycobacterium infection -
XX Example 1; Page 156-157; 161pp; English.
XX The present invention relates to the isolation of Mycobacterium
XX tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids

CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion
 CC of one or more Mycobacterium proteins and nucleic acid molecules
 CC encoding such polypeptides. The Mycobacterium proteins and nucleic acid
 CC molecules encoding them can be used in diagnostic kits for the detection
 CC of Mycobacterium infection in patients and biological samples. The
 CC compounds of the invention and antibodies directed against the
 CC Mycobacterium proteins may be used in vaccines for immunisation against
 CC Mycobacterium infections. The nucleic acids encoding the Mycobacterium
 CC proteins may be used in gene therapy. The present sequence represents
 CC the polypeptide encoded by M. tuberculosis clone H7CC#1.
 XX
 SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 22; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISAIIDGLYDGLGIPNQGGILYSSLEYFEKALELAAAFPDGWLGSAA 207
 Db 1 MSRAFIIDPTISAIIDGLYDGLGIPNQGGILYSSLEYFEKALELAAAFPDGWLGSAA 60

QY 208 DKYAGKRNHNVFFQELADLRQLISLHDQANAVQTTTRDILEGAKKGLFVRPVAVDLT 267
 Db 61 DKYAGKRNHNVFFQELADLRQLISLHDQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAFAFPFCAGAMAVVGGALAVLVKTLINATQLKLAELVAAATAD 327
 Db 121 YIPVVGHALSAFAFPFCAGAMAVVGGALAVLVKTLINATQLKLAELVAAATAD 180

QY 328 IISDVADIIGKLTGEVWEFITNALNGLKELMDKLTGWTGLFSRGSNLEFFAGVPGLT 387
 Db 181 IISDVADIIGKLTGEVWEFITNALNGLKELMDKLTGWTGLFSRGSNLEFFAGVPGLT 240

QY 388 GATSGLSQVTGFGAAGLSASSGLAHADSLASSSLPALAGIGGSGFGGIPSLAQVHAA 447
 Db 241 GATSGLSQVTGFGAAGLSASSGLAHADSLASSSLPALAGIGGSGFGGIPSLAQVHAA 300

QY 448 STRQALRPADGPGVGAAGAEVGGQSQVLSAQSGQGVGVMGMPHSPSSGASKGTTTKY 507
 Db 301 STRQALRPADGPGVGAAGAEVGGQSQVLSAQSGQGVGVMGMPHSPSSGASKGTTTKY 360

QY 508 SEGAAAGTEDAERAPVEADAGGQKVLVRNV 539
 Db 361 SEGAAAGTEDAERAPVEADAGGQKVLVRNV 392

RESULT 5
 AAU01882
 ID AAU01882 standard; Protein; 392 AA.
 XX
 AC AAU01882;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1/Mtb40.
 XX
 KW HTCC#1; Mtb40; antigen; vaccine; tuberculosis; AIDS;
 XX acquired immunodeficiency disease.
 OS Mycobacterium tuberculosis.
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Peptide 1
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 14..33
 FT /label= Peptide 2
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 27..46
 FT /label= Peptide 3
 FT /note= "Used for T-cell epitope mapping"

FT Peptide 40..59
 FT /label= Peptide 4
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 53..72
 FT /label= Peptide 5
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 66..85
 FT /label= Peptide 6
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 79..98
 FT /label= Peptide 7
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 92..111
 FT /label= Peptide 8
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 105..124
 FT /label= Peptide 9
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 118..137
 FT /label= Peptide 10
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 131..150
 FT /label= Peptide 11
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 144..163
 FT /label= Peptide 12
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 157..176
 FT /label= Peptide 13
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 170..189
 FT /label= Peptide 14
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 183..202
 FT /label= Peptide 15
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 196..215
 FT /label= Peptide 16
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 209..228
 FT /label= Peptide 17
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 222..241
 FT /label= Peptide 18
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 235..254
 FT /label= Peptide 19
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 248..267
 FT /label= Peptide 20
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 261..280
 FT /label= Peptide 21
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 274..293
 FT /label= Peptide 22
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 287..306
 FT /label= Peptide 23
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 300..319
 FT /label= Peptide 24
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 313..332
 FT /label= Peptide 25
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 325..344
 FT /label= Peptide 26
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 339..358
 FT /label= Peptide 27
 FT /note= "Used for T-cell epitope mapping"
 FT Peptide 352..371

FT		/label= Peptide28	
FT	Peptide	/note= "Used for T-cell epitope mapping"	
FT		365..384	
FT		/label= Peptide 29	
FT		/note= "Used for T-cell epitope mapping"	
FT	Peptide	373..392	
FT		/label= Peptide 30	
FT		/note= "Used for T-cell epitope mapping"	
XX	WO200124820-A1.		
XX	12-APR-2001.		
XX	10-OCT-2000; 2000WO-US28095.		
XX	07-OCT-1999; 99US-0158338.		
PR	07-OCT-1999; 99US-0158425.		
XX	(CORI-) CORIXA CORP.		
XX	Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;		
PI	WPI; 2001-290576/30.		
DR	N-ESDB; AAS03773.		
XX	Vaccinating against Mycobacteria infections in mammals using fusion proteins comprising combinations of heterologous antigens -		
PT	Example 2; Fig 6; 168pp; English.		
PS	The sequence represents Mycobacterium tuberculosis HTCC#1 (also known as Mtb401), an M. tuberculosis antigen. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by Mycobacterieia. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS.		
CC	Sequence 392 AA;		
QQ	Query Match 71.7%; Score 1963; DB 22; Length 392;		
QQ	Best Local Similarity 99.7%; Pred. No. 3e-155;		
QQ	Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	148 MSRAFIIDPTISADGLYDLIGIGIPNOGGILYSLEYFEKALBELAAAFPGDGLGSA 207		
Db	1 MSRAFIIDPTISADGLYDLIGIGIPNOGGILYSLEYFEKALBELAAAFPGDGLGSA 60		
QY	208 DKVAGKNRNHVNFQELADLDROLISLIHDQANAVQTTRDILEGAKKGLEFVRPVVDLT 267		
Db	61 DKVAGKNRNHVNFQELADLDROLISLIHDQANAVQTTRDILEGAKKGLEFVRPVVDLT 120		
QY	268 YIPVWGHALSAAFOAPFCAGAMAVVGGNALVLVVKTLLINATQLIKLAKLAELVAIAAD 327		
Db	121 YIPVWGHALSAAFOAPFCAGAMAVVGALAVLVVKTLLINATQLIKLAKLAELVAIAAD 180		
QY	328 IISDVADIIRKGIIEGVWEFITNALNGKLWDKLTGWVTGLFSRGWNLESFFFAVPGLT 387		
Db	181 IISDVADIIRKGIIEGVWEFITNALNGKLWDKLTGWVTGLFSRGWNLESFFFAVPGLT 240		
QY	388 GATSGLSQVTGLFGAAGLSASSGLAHADSLSASSSLPALACIGGGSGFGGLPSLAQHAA 447		
Db	241 GATSGLSQVTGLFGAAGLSASSGLAHADSLSASSSLPALACIGGGSGFGGLPSLAQHAA 300		
QY	448 STRQALRPADCPVGAAAEOVGGSQQLVSAGCSOGMGFPVGMGWHPPSSGASKGTITTKY 507		
Db	301 STRQALRPADCPVGAAAEOVGGSQQLVSAGCSOGMGFPVGMGWHPPSSGASKGTITTKY 360		
QY	508 SEGAAAGTEDAERAPVEADAGGGOKVLVRNV 539		
Db	361 SEGAAAGTEDAERAPVEADAGGGOKVLVRNV 392		

Db 181 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 240
 QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 Db 241 GATSGLSQVTLGFGAAGLSAGSLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 448 STQALRPRADGPVGAAGAAQVGGQSLVSAQSQGSGGPGVGMGHPSPSGASKGTTTKY 507
 Db 301 STQALRPRADGPVGAAGAAQVGGQSLVSAQSQGSGGPGVGMGHPSPSGASKGTTTKY 360
 QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
 Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 7

AAE29714
 ID AAE29714 standard; Protein; 392 AA.

AC AAE29714;

XX 27-JAN-2003 (first entry)
 DT
 XX

DE Mycobacterium sp. HTCC#1 antigenic protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; HTCC#1 antigen;
 KW MTB40.

XX Mycobacterium sp.

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US08223.

XX 13-MAR-2001; 2001US-275837P.

XX (CORI-) CORIXA CORP.

PI Skeiky Y, Brannon M, Guderian J;

XX WPI; 2002-759844/G2.

DR N-PSDB; AAD47091.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA,
 PT LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
 PT immunity against pathogenic microorganisms e.g. Leishmania and
 PT Mycobacterium tuberculosis

XX Disclosure; Page 103-104; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides,
 CC as in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Mycobacterium sp. HTCC#1 antigenic protein. HTCC#1 is also referred
 CC to as MTB40.

XX Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 23; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISALDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSSAA 207
 Db 1 MSRAFIIDPTISALDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSSAA 60
 QY 208 DKVAGKRNHNFFQELADLDROLISLIHDQANAVOTTRDILLEGAKKGLFVRPVAVDLT 267
 Db 61 DKVAGKRNHNFFQELADLDROLISLIHDQANAVOTTRDILLEGAKKGLFVRPVAVDLT 120
 QY 268 YIPVGHASAAQAQPCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 327
 Db 121 YIPVGHASAAQAQPCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAD 180
 QY 328 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 387
 Db 181 IISDVADIKGTGGEVWEFTNALNGKELWDLKLTGWVTGLFSRGSNLESFFAGVPGLT 240
 QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
 Db 241 GATSGLSQVTLGFGAAGLSAGSLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 448 STQALRPRADGPVGAAGAAQVGGQSLVSAQSQGSGGPGVGMGHPSPSGASKGTTTKY 507
 Db 301 STQALRPRADGPVGAAGAAQVGGQSLVSAQSQGSGGPGVGMGHPSPSGASKGTTTKY 360
 QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
 Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 8

AAE17578
 ID AAE17578 standard; Protein; 392 AA.

XX AAE17578;

XX 22-APR-2002 (first entry)

XX Mycobacterium species MTB40 (HTCC #1) protein.

XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB40; HTCC #1 protein.

XX Mycobacterium sp.

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US19959.

XX 20-JUN-2000; 2000US-0597796.

XX 01-FEB-2001; 2001US-265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

XX N-PSDB; AAD28349.

XX Composition comprising MTB39 antigen and MTB32A antigen from

XX Mycobacterium species, useful for eliciting immune response in a

XX subject

XX Claim 9; Page 120; 136pp; English.

XX The present invention relates to fusion proteins containing at least
 CC two Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and

CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected
 CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB40 (H7CC #1) protein.

XX SQ Sequence 392 AA;

Query Match 71.7%; Score 1963; DB 23; Length 392;
 Best Local Similarity 99.7%; Pred. No. 3e-155;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 207
 DB 1 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 60
 QY 208 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 267
 DB 61 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
 QY 268 YIPVVGHALSAAPFAFCAGAMVVGALAYLVVVKTLINATQLLKLAELVAIAAD 327
 DB 121 YIPVVGHALSAAPFAFCAGAMVVGALAYLVVVKTLINATQLLKLAELVAIAAD 180
 QY 328 IISVDADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 387
 DB 181 IISVDADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 240
 QY 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQV 447
 DB 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQV 300
 QY 448 STROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 507
 DB 301 STROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 360
 QY 508 SEGAAAGTDAERAPVEADAGGOKVLVRNVV 539
 DB 361 SEGAAAGTDAERAPVEADAGGOKVLVRNVV 392

RESULT 9

ID AAU01886
 XX AAU01886 standard; Protein; 408 AA.
 AC AAU01886;
 XX 29-AUG-2001 (first entry)
 DT M. tuberculosis antigen HTCC#1(TM-1).
 XX HTCC#1(TM-1); antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease; His Tag.
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 FH Binding-site 1..8
 FT /label= Histidine_tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 169..180
 FT /note= "These residues are deleted to form the
 FT HTCC#1(TM-1) protein sequence"

FT Misc-difference 405
 FT /label= OTHER
 FT /note= "In frame STOP codon"

PN WO200124820-A1.

XX 12-APR-2001.

PD 10-OCT-2000; 2000WO-US28095.

XX 07-OCT-1999; 99US-0158338.

PR 07-OCT-1999; 99US-0158425.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;

XX WPI; 2001-290576/30.

DR N-PSDB; AAS03777.

XX Vaccinating against Mycobacteria infections in mammals using fusion

PT proteins comprising combinations of heterologous antigens -

XX Example 2; Fig 9a; 168pp; English.

XX The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
 CC representing HTCC#1 (an M. tuberculosis antigen) lacking the first
 CC transmembrane domain and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous antigens,
 CC as a fusion protein, and vectors expressing the fusion proteins are used
 CC as vaccines to prophylactically immunise mammals (especially humans)
 CC against infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.

XX SQ Sequence 408 AA;

Query Match 70.9%; Score 1941; DB 22; Length 408;
 Best Local Similarity 98.7%; Pred. No. 2.1e-153;
 Matches 391; Conservative 0; Mismatches 1; Indels 4; Gaps 2;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 207

DB 9 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 68

QY 208 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 267

DB 69 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 128

QY 268 YIPVVGHALSAAPFAFCAGAMVVGGA--LAYLVVKTLINA--TQLLKLAELVA 323

DB 129 YIPVVGHALSAAPFAFCAGAMVVGGA--LAYLVVKTLINA--TQLLKLAELVA 188

QY 324 AIADIISVDADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 383

DB 189 AIADIISVDADIIKGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 248

QY 384 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 443

DB 249 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 308

QY 444 VHAASTROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 503

DB 309 VHAASTROALPRADPGPVGAAGAEQVGGOSQLVSAQSGQMGPGVGMGMHPSSGASKGTT 368

QY 504 TKKYSAGAAAGTDAERAPVEADAGGOKVLVRNVV 539

DB 369 TKKYSAGAAAGTDAERAPVEADAGGOKVLVRNVV 404

RESULT 10
 AAU01887

ID AAU01887 standard; Protein; 408 AA.
 AC AAU01887;
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis antigen HTCC#1(TM-2).
 XX HTCC#1(TM-2); antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease; His Tag.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PH Key Location/Qualifiers
 FT Binding-site 1..8
 FT /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 110..213
 FT /note= "These residues are deleted to form the
 FT HTCC#1(TM-2) protein sequence"
 FT Misc-difference 405
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 FT
 XX WO200124820-A1.
 XX
 XX 12-APR-2001.
 XX
 XX 10-OCT-2000; 2000WO-US28095.
 XX
 XX 07-OCT-1999; 99US-0158338.
 XX
 XX 07-OCT-1999; 99US-0158425.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 XX
 XX N-PSDB; AAS03778.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 XX proteins comprising combinations of heterologous antigens -
 XX
 XX Example 2; Fig 9d; 168pp; English.
 XX
 XX The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
 XX representing HTCC#1 (an M. tuberculosis antigen) lacking all the
 XX transmembrane domains and includes a His tag at the N-terminus to aid
 XX purification. Compositions comprising at least 2 heterologous antigens,
 XX as a fusion protein, and vectors expressing the fusion proteins are used
 XX as vaccines to prophylactically immunise mammals (especially humans)
 XX against infection by Mycobacteria. The compositions contain at least 2
 XX heterologous antigens that increase the serological sensitivity of
 XX individuals infected with tuberculosis, a disease frequently affecting
 XX patients with acquired immunodeficiency disease, AIDS.
 XX
 XX Sequence 408 AA;
 SQ
 Query Match 70.9%; Score 1941; DB 22; Length 408;
 Best Local Similarity 98.7%; Pred. No. 2.1e-153;
 Matches 391; Conservative 0; Mismatches 1; Indels 4; Gaps 2;
 QY 148 MSRAFIIDPTISAIIDGLYDILLGIPNCGGILYSLEYFEKALEBLAAFFCDGWLGSAA 207
 DB 9 MSRAFIIDPTISAIIDGLYDILLGIPNCGGILYSLEYFEKALEBLAAFFCDGWLGSAA 68
 QY 208 DKYAGKRVNHNFFQELADLDQLISLHDQANAVQTTDR--ILEGAKKGLFEFVRPVAVD 265
 DB 69 DKYAGKRVNHNFFQELADLDQLISLHDQANAVQTTDRKILLEGAKKGLFEFVRPVAVD 128
 QY 266 LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAELVAAAI 325

DB 129 LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAELVAAAI 188
 QY 326 ADIISDVADIIGKILGEVWEFITNA--LNGKLKELMDKLTGWVTGLFSRGWSNLESFFAGV 383
 DB 189 ADIISDVADIIGKILGEVWEFITNAKLNLGLKELMDKLTGWVTGLFSRGWSNLESFFAGV 248
 QY 384 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 443
 DB 249 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 308
 QY 444 VHAASRQALRPRAADPGVGAARAEQVGGQSLVSAQSQGMGPGVGMGHPSSGASKGTT 503
 DB 309 VHAASRQALRPRAADPGVGAARAEQVGGQSLVSAQSQGMGPGVGMGHPSSGASKGTT 368
 QY 504 TKKYSAGAAAGTDAERAPVEADAGGQKVLVRNVV 539
 DB 369 TKKYSAGAAAGTDAERAPVEADAGGQKVLVRNVV 404
 RESULT 11
 AAU01903
 ID AAU01903 standard; Protein; 788 AA.
 XX
 XX AAU01903;
 XX
 XX 29-AUG-2001 (first entry)
 XX
 XX M. tuberculosis antigen HTCC#1 fusion protein #2.
 XX
 XX HTCC#1(1-149)-TbH9-HTCC#1(161-392); antigen; vaccine; tuberculosis;
 KW AIDS; acquired immunodeficiency disease; His Tag.
 XX
 XX Mycobacterium tuberculosis.
 XX
 XX Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Binding-site 3..8
 FT /label= Histidine tag
 FT /note= "Nickel chelating region used to aid
 FT purification of the protein"
 FT Misc-difference 785
 FT /label= OTHER
 FT /note= "In frame STOP codon"
 FT
 XX WO200124820-A1.
 XX
 XX 12-APR-2001.
 XX
 XX 10-OCT-2000; 2000WO-US28095.
 XX
 XX 07-OCT-1999; 99US-0158338.
 XX
 XX 07-OCT-1999; 99US-0158425.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 XX proteins comprising combinations of heterologous antigens -
 XX
 XX Example 2; Fig 11; 168pp; English.
 XX
 XX The sequence represents Mycobacterium tuberculosis fusion protein,
 XX HTCC#1(1-149)-TbH9-HTCC#1(161-392) and includes a His tag at the
 XX N-terminus to aid purification. Compositions comprising at least 2
 XX heterologous antigens, as a fusion protein, and vectors expressing the
 XX fusion proteins are used as vaccines to prophylactically immunise mammals
 XX (especially humans) against infection by Mycobacteria. The compositions
 XX contain at least 2 heterologous antigens that increase the serological
 XX sensitivity of individuals infected with tuberculosis, a disease
 XX frequently affecting patients with acquired immunodeficiency disease,

```
CC AIDS.
XX
SQ Sequence 788 AA;
Query Match 62.8%; Score 1718; DB 22; Length 788;
Best Local Similarity 49.5%; Pred. No. 2.2e-134;
Matches 384; Conservative 2; Mismatches 6; Indels 384; Gaps 2;
QY 148 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 9 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 68
QY 208 DRYAGKRNHNVFFQELADRLQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 267
DB 69 DRYAGKRNHNVFFQELADRLQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 128
QY 268 YIPVVGHALSAFAQPFACAGAMVVGALAYLV 300
DB 129 YIPVVGHALSAFAQPFACAGAMVVGALAYLV 188
QY 301 300
DB 189 QMDSVADLPSAASAFQSVVWGLTVGWSAGLMVAAPVYVAMSVTAGAELTAA 248
QY 301 300
DB 249 QVRVAABAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPAIVNEAYEGEMWAQDA 308
QY 301 305
DB 309 AAMFGYAAATATATATLIPPEEAPMTSAGLLBQAAAEASDTAAANGLMNNVPOALQ 368
QY 306 305
DB 369 QLAQTOGTPSSKLGGLKWTVSPHRSPISNNWGMANNHSMNTNSGVSWTNTLSMLKGF 428
QY 306 305
DB 429 AAAAAQAQVTAQAQNGVRAMSSLSGLSGGVAANLGRAASVGLSVPOAWAANQ 488
QY 306 305
DB 489 AVTFAARALPUTSITAAREPGQVGLGFLPVQWGARAGGLGVLVPRPRYVYHPSPA 548
QY 306 363
DB 549 AGKLTQLLKLAKLAELVAARAIADIISDVADIISGLGEVWEFTINALNGKELWDKLTG 608
QY 364 WVTGLFSRGWNLSEFPFAGVPEGLTGATSGLSQVTLGFCAGLSASSGLAHADSLSASL 423
DB 609 WVTGLFSRGWNLSEFPFAGVPEGLTGATSGLSQVTLGFCAGLSASSGLAHADSLSASL 668
QY 424 PALAGIGGGGFGGLPSLAQVHAASSTQALRPRADGPVGAAGAAEOVGGSOLVSAQGSQGM 483
DB 669 PALAGIGGGGFGGLPSLAQVHAASSTQALRPRADGPVGAAGAAEOVGGSOLVSAQGSQGM 728
QY 484 GGPVGMGMHPSSGASKGTTTTKKYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 539
DB 729 GGPVGMGMHPSSGASKGTTTTKKYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 784
RESULT 12
ABU05402
ID ABU05402 standard; Protein; 394 AA.
XX
AC ABU05402;
XX
DT 08-APR-2003 (first entry)
XX
DE M. tuberculosis and M. leprae marker protein #53.
XX
KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
XX Mycobacterial disease; tuberculosis; leprosy.
```

```
XX Mycobacterium tuberculosis.
OS Mycobacterium leprae.
XX
PN WO200274903-A2.
XX
PD 26-SEP-2002.
XX
XX 22-FEB-2002; 2002WO-IB01973.
XX
XX 22-FEB-2001; 2001US-270123P.
XX
PA (INSP ) INST PASTEUR.
XX
PI Cole S;
XX
DR WPI; 2002-759885/82.
XX
PT Identifying and selecting genes for survival or virulence of
PT mycobacteria by a comparative genomic analysis of the sequences of
PT Mycobacterium tuberculosis and M. leprae -
XX
XX Claim 17; Page 224-225; 874pp; English.
XX
XX This invention relates to a novel method for identifying essential genes
XX for survival or virulence of mycobacteria species. The method comprises
XX aligning the genomic sequence of a first mycobacterium species on a
XX genomic sequence of a second mycobacterium species and selecting a
XX polynucleotide sequence that is highly conserved in both genomes with no
XX counterparts in other bacterial genomic sequences and that corresponds
XX to an essential gene for the survival or virulence of mycobacterium
XX species. The method of the invention is useful for detecting M.
XX tuberculosis or M. leprae infection. The method reduces the number of
XX potential new targets and protective antigens for new drugs and vaccine
XX compositions to treat and prevent mycobacterial diseases, particularly
XX tuberculosis and leprosy. The present sequence represents a marker
XX protein from Mycobacterium tuberculosis and Mycobacterium leprae
XX identified using the method of the invention.
XX
XX Sequence 394 AA;
Query Match 46.1%; Score 1263; DB 23; Length 394;
Best Local Similarity 62.9%; Pred. No. 7.8e-97;
Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;
QY 148 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 1 MSRAFIIDPTISAIDGLYDLIGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
QY 208 DRYAGKRNHNVFFQELADRLQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 267
DB 61 DRYAGKRNHNVFFQELADRLQLISLIHDOANAVQTTDRILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAFAQPFACAGAMVVGALAYLVVKTLLINATOLLKLAKLAELVAARAIAD 327
DB 121 YIPVVGHALSAFAQPFACAGAMVVGALAYLVVKTLLINATOLLKLAKLAELVAARAIAD 180
QY 328 IISDVADIISGLGEVWEFTINALNGKELWDKLTGVTGVTGVTGVTGVTGVTGVTGVTG 387
DB 181 VVSDGVAIIKIGIVDHLWHEFIAGLTGLKDIVKIIHWFGELFSHNSRSLHSEFGGIPGLS 240
QY 388 GATSGLSQVTLGFCAGLSASSGLAHADSLSASL PALAGIGGGGFGGLPSLAQVHA 447
DB 241 GATSGLSQVTLGFCAGLSASSGLAHADSLSASL PALAGIGGGGFGGLPSLAQVHA 300
QY 448 STRQALRPRADGPVGAAGAAEOVGGSOLVSAQGSQGMVGMGMHPSSGASK -GTTTK 505
DB 301 STRQTRSGAGVSAELSTEQFGQGEFVSAQGSQGMVGMGMHPSSGASK -GTTTK 360
QY 506 KYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 539
DB 361 KYSEGAAGCTEDAERAPVEADAGGQKVLVRNV 394
```

```

RESULT 13
AAU01883
ID AAU01883 standard; Protein; 242 AA.
XX
AC AAU01883;
XX
DT 29-AUG-2001 (first entry)
XX
DE M. tuberculosis antigen HTCC#1(1-223).
XX
KW HTCC#1(1-223); antigen; vaccine; tuberculosis; AIDS;
KW acquired immunodeficiency disease; His Tag.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Binding-site 2..7 /label= Histidine tag
FT /note= "Nickel chelating region used to aid
FT purification of the protein"
FT Misc-difference 240
FT /label= OTHER
FT /note= "In frame STOP codon"
XX
PN WO200124820-A1.
XX
PD 12-APR-2001.
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
PR 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158425.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR WPI; 2001-290576/30.
DR N-PSDB; AAS03774.
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
PS Example 2; Fig 7a; 168pp; English.
XX
CC The sequence is Mycobacterium tuberculosis HTCC#1(1-223),
CC representing the first transmembrane domain of HTCC#1 (an M.
CC tuberculosis antigen) and includes a His tag at the N-terminus to aid
CC purification. Compositions comprising at least 2 heterologous antigens,
CC as a fusion protein, and vectors expressing the fusion proteins are used
CC as vaccines to prophylactically immunise mammals (especially humans)
CC against infection by Mycobacteria. The compositions contain at least 2
CC heterologous antigens that increase the serological sensitivity of
CC individuals infected with tuberculosis, a disease frequently affecting
CC patients with acquired immunodeficiency disease, AIDS.
XX
SQ Sequence 242 AA;
Query Match 42.7%; Score 1168; DB 22; Length 242;
Best Local Similarity 99.6%; Pred. No. 3 3e-89;
Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 148 MSRAFIIDPTISAIIDGLYDILLGIPNCGIILSYSSLEYFEKALEBELAAAFPGDGLGSA 207
DB 8 MSRAFIIDPTISAIIDGLYDILLGIPNCGIILSYSSLEYFEKALEBELAAAFPGDGLGSA 67
OY 208 DKVAGKRNHNVFFQELADRLQLISLHDQANAVQTTDRDILEGAKKGLFVRPVAVDLT 267
DB 68 DKVAGKRNHNVFFQELADRLQLISLHDQANAVQTTDRDILEGAKKGLFVRPVAVDLT 127
OY 268 YIPVVGHALSAFAQAPFCAGAMVGGALAYLVVKTILNATQLLKLAELVAAAIAD 327
DB 9 DVADIINGILGEVWEFITNALNGLKELWDKLTGCVTGLFSRGWSNLESPFAGVGLTGAT 68

Db 128 YIPVVGHALSAFAQAPFCAGAMVGGALAYLVVKTILNATQLLKLAELVAAAIAD 187
OY 328 IISDVADIINGILGEVWEFITNALNGLKELWDKLTGCVTGLFSRGWSNLESPF 379
XX
DB 188 IISDVADIINGILGEVWEFITNALNGLKELWDKLTGCVTGLFSRGWSNLESPF 239

RESULT 14
AAU01902
ID AAU01902 standard; Protein; 744 AA.
XX
AC AAU01902;
XX
DT 29-AUG-2001 (first entry)
XX
DE M. tuberculosis antigen HTCC#1 fusion protein #1.
XX
KW HTCC#1(184-392)-TBH9-HTCC#1(1-129); antigen; vaccine; tuberculosis;
KW AIDS; acquired immunodeficiency disease; His Tag.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Binding-site 3..8 /label= Histidine tag
FT /note= "Nickel chelating region used to aid
FT purification of the protein"
FT Misc-difference 742
FT /label= OTHER
FT /note= "In frame STOP codon"
XX
PN WO200124820-A1.
XX
PD 12-APR-2001.
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
PR 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158425.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR WPI; 2001-290576/30.
DR N-PSDB; AAS03790.
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens -
XX
PS Claim 14; Fig 10; 168pp; English.
XX
CC The sequence represents Mycobacterium tuberculosis fusion protein,
CC HTCC#1(184-392)-TBH9-HTCC#1(1-129) and includes a His tag at the
CC N-terminus to aid purification. Compositions comprising at least 2
CC heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise mammals
CC (especially humans) against infection by Mycobacteria. The compositions
CC contain at least 2 heterologous antigens that increase the serological
CC sensitivity of individuals infected with tuberculosis, a disease
CC frequently affecting patients with acquired immunodeficiency disease,
CC AIDS.
XX
SQ Sequence 744 AA;
Query Match 38.7%; Score 1060; DB 22; Length 744;
Best Local Similarity 100.0%; Pred. No. 1 7e-79;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 331 DVADIINGILGEVWEFITNALNGLKELWDKLTGCVTGLFSRGWSNLESPFAGVGLTGAT 390
DB 9 DVADIINGILGEVWEFITNALNGLKELWDKLTGCVTGLFSRGWSNLESPFAGVGLTGAT 68

```


Qy	391	SGLSQV	TGLFGA	AGLSAS	SGLIAH	DSIASS	SLPALA	GIGGG	SGFG	GLPSLA	OVHAA	STR	450	
Db	69	SGLSQV	TGLFGA	AGLSAS	SGLIAH	DSIASS	SLPALA	GIGGG	SGFG	GLPSLA	OVHAA	STR	128	
Qy	451	QALRRP	ADGP	VGAAAE	QVGG	SQQL	VSAG	SGSG	MG	PPVGM	GMP	SSGAS	KGTTTKKYS	510
Db	129	QALRRP	ADGP	VGAAAE	QVGG	SQQL	VSAG	SGSG	MG	PPVGM	GMP	SSGAS	KGTTTKKYS	188
Qy	511	AAAGTE	DAER	APVEA	DAGG	GQKVL	VRNV	V	539					
Db	189	AAAGTE	DAER	APVEA	DAGG	GQKVL	VRNV	V	217					

RESULT 15

AAU01904
ID AAU01904 standard; Protein: 815 AA.

AA
AC
AAU01904;

29-AUG-2001 (first entry)

DE M. tuberculosis antigen HTCC#1 fusion protein #3.

XX HTCC#1 (184-392)-TbH9-HTCC#1 (1-200); antigen; vaccine; tuberculosis;
KW AIDS; acquired immunodeficiency disease; His tag.
KW

xx	Mycobacterium tuberculosis
OS	Synthetic.

XX	Key	Location/Qualifiers
FH		

	Binding-site	3.8	/label= Histidine tag	/note= "Nickel chelating resin used to aid purification of the protein"
FT				
FT				
PT				
FT				

```

11  Misc-difference 813
12  FT /label= OTHER
13  FT /note= "In frame STOP codon"
14  FT

```

AA
PN
WO200124820-A1.

12-APR-2001.

XX
PF
10-OCT-2000: 2000WO-US28095.

XX
PR 07-OCT-1999: 99US-0158338.

PR 07-OCT-1999; 99US-0158425.
XX

PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;

XX
DR WPI: 2001-290576/30.

xx	Vaccinating against Mycobacteria infections in mammals using fusion
PT	proteins comprising combinations of heterologous antigens -
PT	proteins comprising combinations of heterologous antigens -
xx	

Example 2: Fig 12: 168pp: English:

The sequence represents *Mycobacterium tuberculosis* fusion protein, H7CNC1 (184-392)-TDH9-H7CNC1 (11-200) and includes a His tag at the N-terminus to aid purification. Compositions comprising at least 2 heterologous antigens, as a fusion protein, and vectors expressing the fusion proteins are used as vaccines to prophylactically immunise mammals (especially humans) against infection by *Mycobacterium tuberculosis*. The compositions contain at least 2 heterologous antigens that increase the serological sensitivity of individuals infected with tuberculosis, a disease frequently affecting patients with acquired immunodeficiency disease, AIDS.

AA	Sequence	815-AA;
SO		

Query Match
38.7%; Score 1060; DB 22; Length 815;

[illegible]

Search completed: November 21, 2003, 16:03:21
Job time : 33.6333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:58:31 ; Search time 11.3667 Seconds
(without alignments)
2006.354 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHHTAASDNFQLSQGG.....RAPVEDAGGQKVLVENVV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	DB ID	Description
1	1963	71.7	392 4	US-09-073-009-138 Sequence 138, App
2	809	29.6	729 4	US-09-223-040-2 Sequence 2, Appli
3	737.5	26.9	585 4	US-09-620-412C-337 Sequence 337, App
4	737.5	26.9	585 4	US-09-598-419-337 Sequence 337, App
5	728	26.6	583 4	US-09-620-412C-353 Sequence 353, App
6	728	26.6	583 4	US-09-598-419-353 Sequence 353, App
7	722.5	26.4	518 4	US-09-620-412C-333 Sequence 333, App
8	722.5	26.4	518 4	US-09-598-419-333 Sequence 333, App
9	717.5	26.2	525 4	US-09-556-877-196 Sequence 196, App
10	717.5	26.2	525 4	US-09-620-412C-196 Sequence 196, App
11	717.5	26.2	525 4	US-09-598-419-196 Sequence 196, App
12	709.5	25.9	715 4	US-09-620-412C-321 Sequence 321, App
13	709.5	25.9	715 4	US-09-598-419-321 Sequence 321, App
14	708	25.9	715 4	US-09-620-412C-329 Sequence 329, App
15	708	25.9	715 4	US-09-598-419-329 Sequence 329, App
16	705	25.8	654 4	US-09-620-412C-341 Sequence 341, App
17	705	25.8	654 4	US-09-598-419-341 Sequence 341, App
18	704	25.7	273 4	US-09-736-457-1864 Sequence 1864, Ap
19	704	25.7	314 4	US-09-736-457-1863 Sequence 1863, Ap
20	704	25.7	691 4	US-09-620-412C-313 Sequence 313, App
21	704	25.7	691 4	US-09-598-419-313 Sequence 313, App
22	701.5	25.6	646 4	US-09-620-412C-317 Sequence 317, App
23	701.5	25.6	646 4	US-09-598-419-317 Sequence 317, App
24	699.5	25.6	700 4	US-09-620-412C-345 Sequence 345, App
25	699.5	25.6	700 4	US-09-598-419-345 Sequence 345, App
26	698	25.5	267 4	US-09-643-597-352 Sequence 352, App
27	698	25.5	267 4	US-09-606-421B-352 Sequence 352, App

ALIGNMENTS

RESULT 1

US-09-073-009-138

; Sequence 138, Application US/09073009

; Patent No. 6555653

; GENERAL INFORMATION:

; APPLICANT: Alderson, Mark

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Campos-Neto, Antonio

; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF

; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY

; STREET: 6300 Columbia Center, 701 Fifth Ave.

; CITY: Seattle

; STATE: Washington

; COUNTRY: US

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073,009

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maxi, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.441C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-622-4900

; TELEFAX: 206-682-6031

; INFORMATION FOR SEQ ID NO: 138:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 392 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-073-009-138

Query Match 71.7%; Score 1963; DB 4; Length 392;

Best Local Similarity 99.7%; Pred. No. 4 5e-166;

Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIPTISADGLYDLIGIPNQGGILYSSLEYFEKALELEAAAPPGDWLGSA 207

Db 1 MSRAFIPTISADGLYDLIGIPNQGGILYSSLEYFEKALELEAAAPPGDWLGSA 60

TUBERCULOSIS AND MEN

```

QY 208 DKYAGKRNHVNFFQELADLDROLISLIHDQANAVOTTRDIILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHVNFFQELADLDROLISLIHDQANAVOTTRDIILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVUGHALSAFAQPCAGAMVVGALAYLVVKTLINATOLLKLAELVAAAAD 327
Db 121 YIPVUGHALSAFAQPCAGAMVVGALAYLVVKTLINATOLLKLAELVAAAAD 180

QY 328 IISDVADIIGKILGEVWEFTITNALNGLKELWDLTGWVTGLFSRGSNLESEFFAGVPGLT 387
Db 181 IISDVADIIGKILGEVWEFTITNALNGLKELWDLTGWVTGLFSRGSNLESEFFAGVPGLT 240

QY 388 GATSGLSQVTLGFAAGLSASSGLAHADSLASASLPALAGIGGSGFGLPSLAQVHAA 447
Db 241 GATSGLSQVTLGFAAGLSASSGLAHADSLASASLPALAGIGGSGFGLPSLAQVHAA 300

QY 448 STRQALRPRADGPGVGAAGAEQVGSQVLSAQSOGMGPGVGMGHPSSGASKGTTTKY 507
Db 301 STRQALRPRADGPGVGAAGAEQVGSQVLSAQSOGMGPGVGMGHPSSGASKGTTTKY 360

QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ. ID NOS: 10
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 29.6%; Score 809; DB 4; Length 729;
Best Local Similarity 40.0%; Pred. No. 3.5e-63;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQLSQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAFAASLGISTGDVITAVDGPINSATAMADALNHHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAFAASLGISTGDVITAVDGPINSATAMADALNHHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEPPAEFLVPRGMSRAFIIDPTISAIIDGLYLLGIG-IP 173
Db 121 TKSGGTRTGNVTLAEPPAEFLVPRGMSRAFIIDPTISAIIDGLYLLGIG-IP 173

QY 173 PNOGGILYSSLEFEKALEELAAAPGDCGLGSAADKYGKRNHVNFFQ-----ELAD 226
Db 175 -SVASDLFSAFAQVGVWGLTVG-----SWIGSSAGLMVAASPVAMSVTAGAELTA 229

QY 227 LDRQLISLIHDQANAVOTTRDIILEGAKKGLFVRPVAVDLT--TYIPVVG-----HA 275
Db 230 AQVRVAAAAYETAYGLTVPPVVI--AENRAELMILIATNLLQNTPAIAVNAEYGEWMA 287

```

```

QY 276 LSAAFOAPFCAGAMVVGALAYLVVKTLINATOLLKLAELVAAAAD--IISDVAD 334
Db 288 QDAAMFGYAAATATATATLLPPEEAPEMTSAGGLEQAAAVEASPTAAANQLMNNVPO 347

QY 335 IIKGI-----LGEVWEFTITNALNGLKELWDLTGWVTGLFSRGSNLESEFFAG 382
Db 348 ALQOLAQTOGTTPSSKLGGLWKTVPSP-----ISNVVSMANNHMSMTNSG 396

QY 383 VPGLTGATSGLSQVTLGFAAG-----LSASSGLAHADSLASASLPALAG-----IGGG 432
Db 397 V-SWNTLS--SMKGFAPAARQAQVTAQNGVRAMSSLGSSLGSSGLGGVAAANLGRA 453

QY 433 SGFGGLPSLAQVHAASTRQALRPRADG-----PVGAAAEQVGSQVLSAQSOGMG- 487
Db 454 ASVGSGL-SVPOAMAAA-NQAVTPAARALPLTSLTSAERPGQW-----LGLLPV 501

QY 488 GMGGMHPSSGAS 499
Db 502 GQMGARAGGGLS 513

RESULT 3
US-09-620-412C-337
; Sequence 337, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-337

Query Match 26.9%; Score 737.5; DB 4; Length 585;
Best Local Similarity 38.8%; Pred. No. 5.5e-57;
Matches 231; Conservative 40; Mismatches 174; Indels 151; Gaps 22;

QY 1 MHHHHHTAASDNFQLSQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQSGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 56

QY 61 NNGNGARVQVVGSAFAASLGISTGDVITAVDGPINSATAMADALNHHHPGDVSVTWQ 120
Db 57 NNGNGARVQVVGSAFAASLGISTGDVITAVDGPINSATAMADALNHHHPGDVSVTWQ 116

QY 121 TKSGGTRTGNVTLAEPPAEFLVPRGMSRAFIIDPTISAIIDGLYLLGIG-IP 173
Db 117 TKSGGTRTGNVTLAEPPAEFLVPRGMSRAFIIDPTISAIIDGLYLLGIG-IP 173

QY 174 NOGGILYSSLE---YFEKA-----LEELAAAFPGDGLGSAADKYGKRNH 217
Db 174 VEGGINFODLEETRIKYNKAGTETKKTITLPSLKAQASAGNADAWASSPQSGS----- 228

QY 218 VNFQELADLDROLISLIHDQANAVOTTRDIILEGAKKGLFVRPVAV-----DLTY 268
Db 229 -----ATTVSDSGSDSSGSDS*SETVPVTAAGGGGLYTDKNLSI 267

QY 269 IPVUGHALSAFAQPCAGAMVVGALAYLVVKTLINATOLLKLAELVAAAADI 328
Db 268 TNITG-IIEIANNKATDVGGGAVVKGTL-----TCENSHR-LQFLKNSDQGG----- 314

QY 329 ISDVADIIGKILGEVWEFTITNALNGLKELWDLTGWVT--GLFSRG-----WSNLEST- 379
Db 315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEGGGLFKGTDXALTWTGLDLSFC 363

```

```
QY 380 -----FAGVPLGTGATGSLQVVG-----LFGAAGLSASSG 410
Db 364 LINNTSEKHGGAFVTKESQYTSVDETIPGIT-PVHGETVITGNKSTGGNGGVCTKR 422
QY 411 LAHADSLASSALPALAGTGGGGLPSLAQVHAASRQALRPADCPVGAAGAEQVGG 470
Db 423 LALSNIQSISISGNSAENG-----AHTCPDSFFTD-----TAEQPA 463
QY 471 QSQLVSAQSGMGPGVGMGHPSPGASKGTTTKYSEGAAAGTDAERAPVEAD 526
Db 464 ASAATSTPKS-----APVSTALSTPSSSVSSLLAASSQSPATSNKETQDPNAD 515

RESULT 4
US-09-598-419-337
; Sequence 337, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-598-419-337

Query Match 26.9%; Score 737.5; DB 4; Length 585;
Best Local Similarity 38.8%; Pred. No. 5.5e-57;
Matches 231; Conservative 40; Mismatches 17; Indels 151; Gaps 22;

QY 1 MHHHHHTAASDNFQSQGGQFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGQFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 56

61 NNGNGARVORVVGSAAPASLGISTGVITAVDGPINSATAMADALNGHHHPGDVISTWQ 120
57 NNGNGARVORVVGSAAPASLGISTGVITAVDGPINSATAMADALNGHHHPGDVISTWQ 116

121 TKSOGTGTGNVTLAGPPAEFL-----VPRGMSRAFIIDPTISAIDGLYDLIG-IP 173
117 TKSOGTGTGNVTLAGPPAEFCRYPSHWEPLGDL-----IQSKQSLFNSYKQGGALY 173

174 NOGGILYSSLE-----YFEKA-----LEELAAFPDGLWLSAADKYAGKRNH 217
174 VEGGINFQDLERIRIKYNKAGTFETKITLPSLKAQASAGNADAWASSPQSGG----- 228

218 VNFQBELADLRQLISLIHQANAVQTTDILEGAKKLEFVRPVAV-----DLTY 268
229 -----ATTVSDGSSGSDSTSETVPTAKGGGLYTDKNLSI 267

269 IPVGHASAAFPAPFCAGAMAVGGALAYLVKTLINATQLLKLAKLAEVAAIAADI 328
268 TMITG-IIEIANNKATDVGGGAYVKGTL-----TCENSHR-LQFLKNSDKQGG----- 314

329 ISDVADIILKILGEVWEFTNALNGIKELWDKLTGWVT--GLFSRG-----WSNLESP- 379
315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEGGGLFIKGTDKALMTGLDSCF 363

380 -----FAGVPLGTGATGSLQVVG-----LFGAAGLSASSG 410
364 LINNTSEKHGGAFVTKESQYTSVDETIPGIT-PVHGETVITGNKSTGGNGGVCTKR 422
411 LAHADSLASSALPALAGTGGGGLPSLAQVHAASRQALRPADCPVGAAGAEQVGG 470
423 LALSNIQSISISGNSAENG-----AHTCPDSFFTD-----TAEQPA 463
```

```
QY 471 QSQLVSAQSGMGPGVGMGHPSPGASKGTTTKYSEGAAAGTDAERAPVEAD 526
Db 464 ASAATSTPKS-----APVSTALSTPSSSVSSLLAASSQSPATSNKETQDPNAD 515

RESULT 5
US-09-620-412C-353
; Sequence 353, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 353
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-353

Query Match 26.6%; Score 728; DB 4; Length 583;
Best Local Similarity 37.8%; Pred. No. 3.8e-56;
Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

QY 1 MHHHHHTAASDNFQSQGGQFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGQFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 56

61 NNGNGARVORVVGSAAPASLGISTGVITAVDGPINSATAMADALNGHHHPGDVISTWQ 120
57 NNGNGARVORVVGSAAPASLGISTGVITAVDGPINSATAMADALNGHHHPGDVISTWQ 116

121 TKSOGTGTGNVTLAGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLIG 171
117 TKSOGTGTGNVTLAGPPAEFCRYPSHWRPLDTQVSESPESTPSD-----DVLGKG 168

172 IPNCGGILYSSLEFYFEKALAEELAAFPDGLWLSAADKYAG-----KNRNHVNFFQ 222
169 -----GGI-----YTEKSLTITGITIDFVSNIAIDSGAGVFTKENLCTNTNSLQFLK 218

223 ELADLD-----RQLISLIHQANAVQTTDIL--EGAKKG-----LEFVR 260
219 NSAGQHGGGAYVTQTHSVNTTSETITPPLVGEVIFSENTAKHGGGICTNKLISLNLK 278

261 PVAV-----DLTYIPVVGHALSAAFPAPFCAGAMAVGGALAYLVKTLIN 306
279 TVTLTKNSAKESGGAIFTDLASITPTDTPESSTPSSSPASTPEVVASA----- 327

307 ATOLLKLAKLAEVAAIAADIISDVAD-----IILKILGEVWEFTNALNG 353
328 --KINRFPASTAPAPSLTEASDQTDQTDTSNDSIDVSIENILVAINQNTSAKKG 395

354 LKELWDKLTGWVTGLFSR--GWSNLESPFAGVPLGTGATS-----GLSQVTGL-FGAAGL 405
386 -----GAIYKXKAKLSRINNLE-----LSGNSSDQVGGGLCTESVEFDAIG- 427

406 SASGLAHADSLASSASL-----PALAGTGGGSGGGLPSLAQVHAASRQALRPADGP 460
428 ---SLUSHYNSAAKEGGEVHSKTVTLISNLKSTTFADNTVKAIVESTPEAPEEIPPEGE 484
461 VAAAAE---QVGGQSQLVSAQSGQMGPGVGMGHPSPGASKGTTTKYSEGAAAGTE 516
485 ESTATENPNSTEGSSANTNLEGSQGDATDTGTGVVNNESQDTSDTGNAESGQLQDSIQ 544

517 DAE 519
545 SNE 547
```

RESULT 6

US-09-598-419-353

; Sequence 353, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C6

; CURRENT APPLICATION NUMBER: US/09/598,419

; CURRENT FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 357

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 353

; LENGTH: 583

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-598-419-353

Query Match

Best Local Similarity 26.6%; Score 728; DB 4; Length 583;

Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIRSGGSPVHIGTAFGLGVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIK-----LPTVHIGTAFGLGVD 56
 QY 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHPGDVSVTWQ 120
 DB 57 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHPGDVSVTWQ 116
 QY 121 TKSOGTRTGNVTLAEGPPABF-----LVPRGSMRAFIIDPTISAIIDGLYLLGI 171
 DB 117 TKSOGTRTGNVTLAEGPPABF-----LVPRGSMRAFIIDPTISAIIDGLYLLGI 168
 QY 172 IPNQGGILYSSLEFYEKALEELAAAFPGDGLWLG-----KNNRNVNFPQ 222
 DB 169 -----GGI-----YTEKSLITIGTIDFVSNIAIDSGAVTKENLSCTWNSIQFLK 218
 QY 223 ELADLD-----ROLISLHDOANAVQT---TRDIL--EGAKG-----LEFVR 260
 DB 219 NSAGQGGGAYVTQWSVNTTSESITPLVGEVIFSENTAKHGCGGICTNKLSSLNLK 278
 QY 261 PVAV-----DUTYIPVVGHALSAAFQAPFCAGAMAVVGGALAVLVKTLIN 306
 DB 279 TVLTKNKSAKESGAIFTDLASITPTDTPESSTPSSSPASTPEVVASA----- 327
 QY 307 ATOLLKLAKLAELVAAAADIISDVAD-----IIGILGEVWEFITNALNG 353
 DB 328 --KINRPFASABPAAPSLTEAESDQTDQETSDTNSDIDVSIENILVAINQNTSAKGG 385
 QY 354 LKELWDLKLTGVTGLFSR--GWSNLSFFAGVPGLTGATS-----CLSQVTGL--FGAAGL 405
 DB 386 -----GAIYKXKAKSLRINWLE-----LSGNSQDVGGGLCTESVEPDIG-- 427
 QY 406 SASGLAHADSLASSASL-----PALAGIGGGGFGGLPSLAQVHAASRQALRPADGP 460
 DB 428 ---SLLSHYNSAAKEGGVIHSKTVTLNLSKSTFTFADNTVKAIVESTPEAPEIIPVEGE 484
 QY 461 VGRAAE-----QVGGQSLVSAQSQGGPGVGMGHPSPSGASKGTTTKYSEGAAGTE 516
 DB 485 ESTATENPNNSWTEGSSANTNLEGSQDGTADTGTGVVNNESQDTSDTGNAESGEQLQDSQ 544
 QY 517 DAE 519
 DB 545 SNE 547

RESULT 7

US-09-620-412C-333

; Sequence 333, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Flind

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 333

; LENGTH: 518

; TYPE: PRT

; ORGANISM: Chlamydia trachomatis

US-09-620-412C-333

Query Match

Best Local Similarity 26.4%; Score 722.5; DB 4; Length 518;

Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIRSGGSPVHIGTAFGLGVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMATAGQIK-----LPTVHIGTAFGLGVD 56
 QY 61 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHPGDVSVTWQ 120
 DB 57 NNGNGARVQVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHPGDVSVTWQ 116
 QY 121 TKSOGTRTGNVTLAEGPPABF-----LVPRGSMRAFIIDPTISAIIDGLYLLGI 170
 DB 117 TKSOGTRTGNVTLAEGPPABF-----LVPRGSMRAFIIDPTISAIIDGLYLLGI 167
 QY 171 GIPNQGGILYSSLEFYEKALEELAAAFPGDGLWLG-----SAAKYAGKNNHV 218
 DB 168 CFPPEKELNFRSVETSSSTFTTETIGAGAGYIVSGNASFTKTNITPTTTTPTNSNS 227
 QY 219 NFFOELADLDRLQLISLHDOANAVQTTTRDILEGAKGLEFVRPVAVDLTYPVVGHALSA 278
 DB 228 SSSGETAS-----VSESDSSTTTTPD-----PKG----- 251
 QY 279 AFQAPFCAGAMAVVGGALAVLVKTLINATQLLKLAKLAELVAAAADIISDVADIIG 338
 DB 252 -----GGA-----FYNH----- 261
 QY 339 ILGEVWEFITNALNGLKELMDKLTGVTGLFSGWSNLSFFAGVPGLTGATSLQVVG 398
 DB 262 VLSFWTRSGTGLSLTSEI--KMTGEGGAIFSGQ-----ELLFTDLTSLT-IQNLSQLSG 314
 QY 399 --LFGAAGLSASSGLAHADSLASSASLPAALAGIGGGGFGGLPSLAQVHAASRQALRP 456
 DB 315 GAIFGGSTISL-SGITKATFSCNSAEVP-----APVKKPTPEK 351
 QY 457 ADGPVGAAGAEVGGQSLVSAQSQGGPGVGMGHPSPSGASKGTTTKYSE-----G 510
 DB 352 AQ-----TASSETS-----SSGNSDVS-----PSSRAEPAAALQSHFICATAT 395
 QY 511 AAAGTDAERAPVEADAGG 530
 DB 396 PAAQTDSTSTPSHKPGSGG 415

RESULT 8

US-09-598-419-333

; Sequence 333, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C6

; CURRENT APPLICATION NUMBER: US/09/598,419

; CURRENT FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 357

```
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 333
; LENGTH: 518
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis
US-09-598-419-333

Query Match      26.4%; Score 722.5; DB 4; Length 518;
Best Local Similarity 37.3%; Pred. No. 9.9e-56;
Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 120
Db 57 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 179
Db 117 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 157
Qy 180 YSSLEYFEKALEELAAAFPGDGLGSAAD 208
Db 158 -----IDGTWEGASGD 169

RESULT 10
US-09-620-412C-196
; Sequence 196, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PR1
; ORGANISM: Chlamydia
US-09-620-412C-196

Query Match      26.2%; Score 717.5; DB 4; Length 525;
Best Local Similarity 71.3%; Pred. No. 2.8e-55;
Matches 149; Conservative 6; Mismatches 13; Indels 41; Gaps 4;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 120
Db 57 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 179
Db 117 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 157
Qy 180 YSSLEYFEKALEELAAAFPGDGLGSAAD 208
Db 158 -----IDGTWEGASGD 169

RESULT 11
US-09-598-419-196
; Sequence 196, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
```

```
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 333
; LENGTH: 518
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis
US-09-598-419-333

Query Match      26.4%; Score 722.5; DB 4; Length 518;
Best Local Similarity 37.3%; Pred. No. 9.9e-56;
Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 120
Db 57 NNGNGARVQVVGSAPAAASLGISTGCVITAVDGPINSATAMADALNGHHHPGDISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 179
Db 117 TKSGGTRTGNVTLAEGPPAEF-LVPRGSMRAF-IIDPTISADGLYDLGI 157
Qy 180 YSSLEYFEKALEELAAAFPGDGLGSAAD 208
Db 158 -----IDGTWEGASGD 169

RESULT 9
US-09-598-419-196
; Sequence 196, Application US/09556877
; Patent No. 6432916
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir
; APPLICANT: Fling, Steve
; APPLICANT: Maisonneuve, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C5
; CURRENT APPLICATION NUMBER: US/09/556,877
; CURRENT FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PR1
; ORGANISM: Chlamydia
```

```
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 196
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-598-419-196

Query Match      26.2%; Score 717.5; DB 4; Length 525;
Best Local Similarity 71.3%; Pred. No. 2.8e-55;
Matches 149; Conservative 6; Mismatches 13; Indels 41; Gaps 4;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-LVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGIL 179
Db 117 TKSGGTRTGNVTLAEGPPAEF-LVPRGMSRAFIIDPTISAIDGLYDLIGIPNQGIL 157
Qy 180 YSSLEVEFEKALEBAAAFDGDHGLGSAAD 208
Db 158 -----IDGTMWEGASGD 169

RESULT 12
US-09-620-412C-321
; Sequence 321, Application US/09620412C
; Patent No. 648234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 321
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-321

Query Match      25.9%; Score 709.5; DB 4; Length 715;
Best Local Similarity 37.5%; Pred. No. 2.3e-54;
Matches 214; Conservative 53; Mismatches 187; Indels 117; Gaps 21;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-----LVPRGMSRA--FIIDPTISAIDGLYDLIG 169
Db 117 TKSGGTRTGNVTLAEGPPAEFCRYPHWRPLDIRTLGKEHNYIKEAPTTLKFGT----- 171
Qy 170 IGIPNQGILYSSLEVEFEKALEBAAAFDGDHGLGSAADKYAGKRNHNVFQELADLR 229
Db 172 LAIEDDAELEIFNIPTQNTSLLA-----LGSATLTVGK----- 207
Qy 121 TKSGGTRTGNVTLAEGPPAEF-----LVPRGMSRA--FIIDPTISAIDGLYDLIG 169
Db 117 TKSGGTRTGNVTLAEGPPAEFCRYPHWRPLDIRTLGKEHNYIKEAPTTLKFGT----- 171
Qy 170 IGIPNQGILYSSLEVEFEKALEBAAAFDGDHGLGSAADKYAGKRNHNVFQELADLR 229
Db 172 LAIEDDAELEIFNIPTQNTSLLA-----LGSATLTVGK----- 207
Qy 230 QLISLIHQANAVQ--TTDRILEGAKKGLFVRPVAVDLTYIPVVGHALS--AAFOAPFC 285
Db 208 -----HGKLNITNLGVILPIILKEGSPPCIRVNPQDMTQNTGTGTPSSSTSTP-- 259
Qy 286 AGAMAVVGGALAYLVKTLINATQLLAKLAEVAAAIADIISDVADIIGKILGEVWE 345
Db 260 ---MIENGRLS-IVDENYESVYDSMDLSRGKAE-----QLISLIETNDGQDSNWQ 308
Qy 346 FITNALN-----GLKELWDKLTGWVTGLPSRGWSNLESFFAGVPGITGATSGLSQV 396
Db 309 ---SSLNTLSLSPPHYGYQGLW--TPNWITTTITLNNNS-----APTSATSAEQ-- 356

Query Match      25.9%; Score 709.5; DB 4; Length 715;
Best Local Similarity 37.5%; Pred. No. 2.3e-54;
Matches 214; Conservative 53; Mismatches 187; Indels 117; Gaps 21;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 56
Qy 61 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISVTWQ 116
Qy 121 TKSGGTRTGNVTLAEGPPAEF-----LVPRGMSRA--FIIDPTISAIDGLYDLIG 169
Db 117 TKSGGTRTGNVTLAEGPPAEFCRYPHWRPLDIRTLGKEHNYIKEAPTTLKFGT----- 171
Qy 170 IGIPNQGILYSSLEVEFEKALEBAAAFDGDHGLGSAADKYAGKRNHNVFQELADLR 229
Db 172 LAIEDDAELEIFNIPTQNTSLLA-----LGSATLTVGK----- 207
Qy 230 QLISLIHQANAVQ--TTDRILEGAKKGLFVRPVAVDLTYIPVVGHALS--AAFOAPFC 285
Db 208 -----HGKLNITNLGVILPIILKEGSPPCIRVNPQDMTQNTGTGTPSSSTSTP-- 259
Qy 286 AGAMAVVGGALAYLVKTLINATQLLAKLAEVAAAIADIISDVADIIGKILGEVWE 345
Db 260 ---MIENGRLS-IVDENYESVYDSMDLSRGKAE-----QLISLIETNDGQDSNWQ 308
Qy 346 FITNALN-----GLKELWDKLTGWVTGLPSRGWSNLESFFAGVPGITGATSGLSQV 396
Db 309 ---SSLNTLSLSPPHYGYQGLW--TPNWITTTITLNNNS-----APTSATSAEQ-- 356
```


QY 397 TGLFGAGLSASGLAHADSLASSALP---ALAGIGGGGFGGLPSLAQVHAASTROAL 453
Db 357 -----KKTSETFTPSNNTTASIPNIKASAGSGSASNSGEVITTKHTLVNWA- 405
QY 454 RPRADGPVGAAGAEQVGGQQLYSAGSQSGQGGPVGWGMHPSS-----GASKGTTTKY 507
Db 406 -----PVGVIYDPIPRGLIANSLVHSGRNMTWGLRLLPDNSWFAALQGAATTLFTKQ 459
QY 508 SEGAAAGTDEARA-PVEADAGG--GQKVLV 535
Db 460 KRLSVHGYSSASKGYTVSSQASGAGHKFLL 490

RESULT 14

US-09-620-412C-329
; Sequence 329, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 329
; LENGTH: 715
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-329

Query Match 25.9%; Score 708; DB 4; Length 715;
Best Local Similarity 70.7%; Pred. No. 3.1e-54;
Matches 152; Conservative 13; Mismatches 26; Indels 24; Gaps 6;
QY 1 MHHHHHTAASDNFOLSGQGGFAIPGQAMAIAGQIRSGGSPTVHIGPTAFGLGLVVD 60
Db 1 MHHHHHTAASDNFOLSGQGGFAIPGQAMAIAGQIK-----LPTVHIGPTAFGLGLVVD 56
QY 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGPINSATAMADALNGHHHPGDIVSVTWQ 116
QY 121 TKSQGTGTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLL 168
Db 117 TKSQGTGTGNVTLAEGPPAEFCRYPHWRPLDPVVQNSAAGASTPSSSSMPG---AV 173
QY 169 GIGIPNQGILYS--SLEYFEKALEEL--AAAPFG 199
Db 174 TINQSGNGSVIFTAESLTPSEK-LQVLNSTSNFPG 207

RESULT 15

US-09-598-419-329
; Sequence 329, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 329
; LENGTH: 715
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis
US-09-598-419-329

Query Match 25.9%; Score 708; DB 4; Length 715;
Best Local Similarity 70.7%; Pred. No. 3.1e-54;
Matches 152; Conservative 13; Mismatches 26; Indels 24; Gaps 6;
QY 1 MHHHHHTAASDNFOLSGQGGFAIPGQAMAIAGQIRSGGSPTVHIGPTAFGLGLVVD 60
Db 1 MHHHHHTAASDNFOLSGQGGFAIPGQAMAIAGQIK-----LPTVHIGPTAFGLGLVVD 56
QY 61 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
Db 57 NNGNGARVQVVGSAAPASLGISTGDIVTAVDGPINSATAMADALNGHHHPGDIVSVTWQ 116
QY 121 TKSQGTGTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLL 168
Db 117 TKSQGTGTGNVTLAEGPPAEFCRYPHWRPLDPVVQNSAAGASTPSSSSMPG---AV 173
QY 169 GIGIPNQGILYS--SLEYFEKALEEL--AAAPFG 199
Db 174 TINQSGNGSVIFTAESLTPSEK-LQVLNSTSNFPG 207

Search completed: November 21, 2003, 16:11:44
Job time : 12.3667 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 16:08:22 ; Search time 21.0222 Seconds
(without alignments)
4680.740 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHHTAAADNFQLSQGG.....RAPVEADAGGQKVLVENVV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1963	71.7	392	9	US-09-073-009-138
2	1963	71.7	392	9	US-09-793-306-138
3	1963	71.7	392	12	US-10-098-732A-29
4	1963	71.7	392	16	US-10-080-170-610
5	1263	46.1	394	16	US-10-080-170-53
6	809	29.6	729	9	US-09-287-849-2
7	809	29.6	729	12	US-10-359-460-2
8	809	29.6	729	12	US-10-098-732A-16
9	809	29.6	729	12	US-10-098-732A-18
10	809	29.6	930	12	US-10-098-732A-65
11	739	27.0	231	9	US-09-287-849-28
12	739	27.0	231	12	US-10-359-460-28
13	737.5	26.9	585	9	US-09-841-132-337
14	728	26.6	583	9	US-09-841-132-353
15	722.5	26.4	518	9	US-09-841-132-333

16	718.5	26.3	529	9	US-09-810-936-324
17	718.5	26.3	529	10	US-09-924-400-324
18	718.5	26.3	529	15	US-10-213-679-324
19	717.5	26.2	525	9	US-09-841-132-196
20	709.5	25.9	715	9	US-09-841-132-321
21	708	25.9	715	9	US-09-841-132-329
22	706.5	25.8	384	10	US-09-902-941-1876
23	706.5	25.8	384	10	US-09-849-626-1876
24	706.5	25.8	384	12	US-10-113-872-1876
25	706.5	25.8	384	15	US-10-017-754-1876
26	705.5	25.8	344	9	US-09-922-217-1085
27	705.5	25.8	344	10	US-09-833-263-1085
28	705.5	25.8	344	14	US-10-025-380-1085
29	705	25.8	654	9	US-09-841-132-341
30	704	25.7	273	10	US-09-736-457-1864
31	704	25.7	273	10	US-09-902-941-1864
32	704	25.7	273	10	US-09-849-626-1864
33	704	25.7	273	12	US-10-113-872-1864
34	704	25.7	273	15	US-10-017-754-1864
35	704	25.7	314	10	US-09-736-457-1863
36	704	25.7	314	10	US-09-902-941-1863
37	704	25.7	314	10	US-09-849-626-1863
38	704	25.7	314	12	US-10-113-872-1863
39	704	25.7	314	15	US-10-017-754-1863
40	704	25.7	400	9	US-09-759-143-852
41	704	25.7	400	9	US-09-780-669-852
42	704	25.7	400	9	US-09-822-827-852
43	704	25.7	400	10	US-09-895-793-852
44	704	25.7	400	10	US-09-895-814-852
45	704	25.7	400	12	US-10-144-678A-852

ALIGNMENTS

RESULT 1

US-09-073-009-138
; Sequence 138, Application US/09073009
; Patent No. US20010012888A1
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid

TUBERCULOSIS AND ME

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-306-138

Query Match
Best Local Similarity 99.7%; Pred. No. 4.9e-145;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

Qy 208 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 267
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

Qy 268 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 327
Db 61 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

Qy 328 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 387
Db 181 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 268 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 327

Qy 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507
Db 121 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 180

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 328 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 387

Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 181 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300

Qy 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507
Db 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507

Qy 301 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 360
Db 301 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 360

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539

Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 3
US-10-098-732A-29
; Sequence 29, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-0120100S
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB40 (HTCC#1)
US-10-098-732A-29

Query Match 71.7%; Score 1963; DB 12; Length 392;
Best Local Similarity 99.7%; Pred. No. 4.9e-145;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

Qy 208 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

Qy 268 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 327
Db 268 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 327

Qy 328 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 387
Db 121 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 180

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 328 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 387

Qy 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507
Db 181 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207

Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 208 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 267

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 61 DKYAGKRNHNHVFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

Qy 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507
Db 268 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 327

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 121 YIPVVGHALSAAPQAPFCAGAMAVVGGALAYLVVKTLLINATQLLLAKLAEVAAAIAD 180

Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 328 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 387

Qy 388 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 447
Db 181 IISDVADIIGKILGEVWEFITTALNGKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy 448 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 507
Db 301 STQALRPADGPGVGAARQVGGOSQOLVSAQSGQGMGPGVGMGMHPSSGASKGTTTKY 360

Qy 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539
Db 508 SEGAAAGTDAERAPVEADAGGQKVLVRNV 539

Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2
US-09-793-306-138
; Sequence 138, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 138
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: hTcc#1
US-09-793-306-138

Query Match 71.7%; Score 1963; DB 9; Length 392;
Best Local Similarity 99.7%; Pred. No. 4.9e-145;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 148 MSRAFIIDPTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
```

```

QY 388 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 447
Db 241 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 300
QY 448 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 507
Db 301 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 4
US-10-080-170-610
; Sequence 610, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; PRIOR FILING DATE: 2002-06-10
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 610
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-610

Query Match 71.7%; Score 1963; DB 16; Length 392;
Best Local Similarity 99.7%; Pred. No. 4.9e-145;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 207
Db 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 60
QY 208 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTTDRILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTTDRILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 180
QY 328 IISDVADIKIGLGEVWEFFITNALNGKELMDKLTGWVTGLFSRGNLSFFPAGVPLT 387
Db 181 IISDVADIKIGLGEVWEFFITNALNGKELMDKLTGWVTGLFSRGNLSFFPAGVPLT 240
QY 388 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 447
Db 241 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 300
QY 448 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 507
Db 301 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 360
QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 5
US-10-080-170-53
; Sequence 53, Application US/10080170
; Publication No. US20030129601A1

```

GENERAL INFORMATION:

```

; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-53

```

Query Match 46.1%; Score 1263; DB 16; Length 394;

Best Local Similarity 62.9%; Pred. No. 2.1e-90;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

```

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 207
Db 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPGDMGWSAA 60
QY 208 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTTDRILEGAKKGLFVRPVAVDLT 267
Db 61 DKYAGKRNHNVFFQELADLRQLISLIHQANAVQTTDRILEGAKKGLFVRPVAVDLT 120
QY 268 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 327
Db 121 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 180
QY 328 IISDVADIKIGLGEVWEFFITNALNGKELMDKLTGWVTGLFSRGNLSFFPAGVPLT 387
Db 181 IISDVADIKIGLGEVWEFFITNALNGKELMDKLTGWVTGLFSRGNLSFFPAGVPLT 240
QY 388 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 447
Db 241 GATGSLQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPLSLAQVHAA 300
QY 448 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 505
Db 301 STQALRPRADPGVGAARAEQVGGOSQLVSAOGSGMGPGVGMGHPSSGASKGTTTKY 360
QY 508 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 539
Db 361 YIPVVGHALSAAPFQAPFCAGAMAVVGGALAYLVVKTLLNATQLKLAKLAELVAAAIAD 394

```

RESULT 6

```

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020003459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18

```

```
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match      29.6%; Score 809; DB 9; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQISGQGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQISGQGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAFPAASLIGSTGDIVITAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
Db 61 NNGNGARVQVVGSAFPAASLIGSTGDIVITAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLGIGI 172
Db 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLGIGI 172
QY 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKRNHVNFPQ-----ELAD 226
Db 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKRNHVNFPQ-----ELAD 226
QY 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGAELTA 229
Db 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGAELTA 229
QY 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
Db 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
QY 230 AQVRVAAAAYETAYGLTVPPVI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
Db 230 AQVRVAAAAYETAYGLTVPPVI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
QY 276 LSAAFQAPFCAGAVVVGALAYLVVKTLINATQLLKLAKLAELVAAAIAAD--IISDVAD 334
Db 276 LSAAFQAPFCAGAVVVGALAYLVVKTLINATQLLKLAKLAELVAAAIAAD--IISDVAD 334
QY 288 QDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNVVQ 347
Db 288 QDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNVVQ 347
QY 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAG 382
Db 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAG 382
QY 348 ALQOLAQPTGTTTPSSKLGGLWKTVPSPRSP-----ISNMVSMANNHMTNSG 396
Db 348 ALQOLAQPTGTTTPSSKLGGLWKTVPSPRSP-----ISNMVSMANNHMTNSG 396
QY 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
Db 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
QY 397 V-SMTNTLS--SMLKGFAPAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
Db 397 V-SMTNTLS--SMLKGFAPAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
QY 433 SGFGGLPSLAQVHAASSTRQALRPADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
Db 433 SGFGGLPSLAQVHAASSTRQALRPADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
QY 454 ASVGSLS-SVPQAWAAA-NQAVTPAARALPLTSLTSAERPGQM-----LGLPLV 501
Db 454 ASVGSLS-SVPQAWAAA-NQAVTPAARALPLTSLTSAERPGQM-----LGLPLV 501
QY 488 GMGGMHPSSGAS 499
Db 502 GQMGARAGGGLS 513

RESULT 7
US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
```

```
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match      29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQISGQGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQISGQGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAFPAASLIGSTGDIVITAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
Db 61 NNGNGARVQVVGSAFPAASLIGSTGDIVITAVDGPINSATAMADALNGHHHPGDIVSVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLGIGI 172
Db 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYDLGIGI 172
QY 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKRNHVNFPQ-----ELAD 226
Db 173 PNQGGILYSSLEFEXKALEELAAAFPCDGLGSAADKYAGKRNHVNFPQ-----ELAD 226
QY 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGAELTA 229
Db 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSSAGLMVAASPYVAMSVTAGAELTA 229
QY 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
Db 227 LDRQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDL--TYIPVVG-----HA 275
QY 230 AQVRVAAAAYETAYGLTVPPVI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
Db 230 AQVRVAAAAYETAYGLTVPPVI--AENRAELMILITATLLGQNTPAIAVNEAEYGEWMA 287
QY 276 LSAAFQAPFCAGAVVVGALAYLVVKTLINATQLLKLAKLAELVAAAIAAD--IISDVAD 334
Db 276 LSAAFQAPFCAGAVVVGALAYLVVKTLINATQLLKLAKLAELVAAAIAAD--IISDVAD 334
QY 288 QDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNVVQ 347
Db 288 QDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNVVQ 347
QY 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAG 382
Db 335 IIKGI-----LGEVWEFITNALNGLKELMDKLTGWVTGLFSRGSNLESFFAG 382
QY 348 ALQOLAQPTGTTTPSSKLGGLWKTVPSPRSP-----ISNMVSMANNHMTNSG 396
Db 348 ALQOLAQPTGTTTPSSKLGGLWKTVPSPRSP-----ISNMVSMANNHMTNSG 396
QY 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
Db 383 VPGLTGATSGLSQVTLGFGAAG-----LSASSGLAHADSLASSASLPALAG-----IGGG 432
QY 397 V-SMTNTLS--SMLKGFAPAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
Db 397 V-SMTNTLS--SMLKGFAPAARQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRA 453
QY 433 SGFGGLPSLAQVHAASSTRQALRPADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
Db 433 SGFGGLPSLAQVHAASSTRQALRPADG-----PVGAAAEQVGGQSGLVSAQSGQMGG-PV 487
QY 454 ASVGSLS-SVPQAWAAA-NQAVTPAARALPLTSLTSAERPGQM-----LGLPLV 501
Db 454 ASVGSLS-SVPQAWAAA-NQAVTPAARALPLTSLTSAERPGQM-----LGLPLV 501
QY 488 GMGGMHPSSGAS 499
Db 502 GQMGARAGGGLS 513

RESULT 8
US-10-098-732A-16
; Sequence 16, Application US/10098732A
; Publication No. US20030175294A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: Protein MTB72F (Ra12-TDH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match      29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 40.0%; Pred. No. 1.4e-54;
Matches 221; Conservative 69; Mismatches 170; Indels 92; Gaps 20;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60

QY 61 NNGNGARVQVVGSAPASLSGISTGDVITAVDGCAPINSATAMADALNGHFGDVISVTWQ 120
DB 61 NNGNGARVQVVGSAPASLSGISTGDVITAVDGCAPINSATAMADALNGHFGDVISVTWQ 120

QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYLLGIGI 172
DB 121 TKSGGTRTGNVTLAGPPAEFVDFGALPPEINSARMYAGSGASLVAAAQMW----- 174

QY 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKRNHNHVFQ-----ELAD 226
DB 173 -SVASDLFSAASAFQSVWGLTVG-----SWIGSSAGLMVAASPYVAMMSVTAGQELTA 229

QY 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRVAVDL--TYIPVVG-----HA 275
DB 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILIATNLGQNTPAIAYNEAEGEMWA 287

QY 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAELVAAAIAD--IISDVAD 334
DB 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVQ 347

QY 335 IIKGI-----LGEVWEFITNALNGKLKELMDKLTGWVTGLFSRGWSNLESFFAG 382
DB 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSP-----ISNMVSMANNHSMNTNSG 396

QY 383 VPGLTGATSGLSQVTLGFGAAGLSASSGLAHADSLSASSASLPALAG-----ICGG 432
DB 397 V-SMTNTLS--SMLKGPAPAAARQAQVTAQNGVYRAMSSLSGLSGSLGGGVAAIIGRA 453

QY 433 SGFGCLPSLAQVHAASTROALRPADG-----PVGAAAEQVGGQQLVSAQSGQMG- 487
DB 454 ASVSSL-SVPOWAAA--NOAVTPAARALPLTSLTSAARGFQGM-----LGGLPV 501

QY 488 QMGGMHPSSSGAS 499
DB 502 QMGARAGGGLS 513

```

RESULT 9
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FmutSA
; OTHER INFORMATION: (Ra12-TDH9-Ra35MutSA)
US-10-098-732A-18

```

```

Query Match      29.6%; Score 809; DB 12; Length 729;
Best Local Similarity 39.5%; Pred. No. 1.4e-54;
Matches 220; Conservative 66; Mismatches 169; Indels 102; Gaps 18;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60

QY 61 NNGNGARVQVVGSAPASLSGISTGDVITAVDGCAPINSATAMADALNGHFGDVISVTWQ 120
DB 61 NNGNGARVQVVGSAPASLSGISTGDVITAVDGCAPINSATAMADALNGHFGDVISVTWQ 120

QY 121 TKSGGTRTGNVTLAGPPAEFLVPRGSM-----SRAFIIDP---TISAIDGLYLLGIGI 172
DB 121 TKSGGTRTGNVTLAGPPAEFVDFGALPPEINSARMYAGSGASLVAAAQMW----- 174

QY 173 PNOGGILYSSLEYFEKALEELAAAFPGDGLGSAADKYAGKRNHNHVFQ-----ELAD 226
DB 175 -SVASDLFSAASAFQSVWGLTVG-----SWIGSSAGLMVAASPYVAMMSVTAGQELTA 229

QY 227 LDRQLISLIHQANAVQTTDRDILEGAKKGLFVRVAVDL--TYIPVVG-----HA 275
DB 230 AQVRVAAAAYETAYGLTVPPVPI--AENRAELMILIATNLGQNTPAIAYNEAEGEMWA 287

QY 276 LSAFQAPFCAGAMAVVGGALAYLVVKTILINATQLKLLAKLAELVAAAIAD--IISDVAD 334
DB 288 QDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNVQ 347

QY 335 IIKGI-----LGEVWEFITNALNGKLKELMDKLTGWVTGLFSRGWSNLESFFAG 382
DB 348 ALQQLAQPTQGTTPSSKGLGKLVSPHRSP-----SNMVSMAWN 388

QY 383 VPGLTGATSGLSQVTLGFGAAGLSASSGLAHADSLSASSASLPALAG-----CGSGFGG 437
DB 389 HNSMT--NGVSMNTLSNMLKGFAPAAQAQVTAQNGVYRAMSSLSGLSGSLGGGV 446

QY 438 -----LPSLAQVHA--ASTROALRPADG-----PVGAAAEQVGGQQLVSAQSGQMG 483
DB 447 AANLGRAASVGSLSVPOWAAA--NOAVTPAARALPLTSLTSAARGFQGM-----L 496

QY 484 GG-PVCGMGHPSSSGAS 499
DB 497 GGLPVQMGARAGGGLS 513

```

RESULT 10
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:

APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
PRIOR FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 55
LENGTH: 930
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
OTHER INFORMATION: (r95f) fusion construct TB MTB72F (Ra12-TbH9-Ra35)
OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 29.6%; Score 809; DB 12; Length 930;
Best Local Similarity 39.5%; Pred. No. 1.9e-54;
Matches 220; Conservative 66; Mismatches 169; Indels 102; Gaps 18;

QY 1 MHHHHHTAASDNFQSQGGGFAIPGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGFAIPGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARVQRVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQRVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISADGLYDLLIGTI 172
DB 121 TKSGGTRTGNVTLAEGPPAEFLVPRGSM-----SRAFIIDP---TISADGLYDLLIGTI 174

QY 173 PNOGGTLYSSLEYFEKALEELAAFPDGLGSAADKYAGKRNHNVFFQ-----ELAD 226
DB 175 -SVASDLFSAASAFQSVVWGLTVG-----SWIGSAGLMVAASPYVAMSVTAGQAEITA 229

QY 227 LDRQLSLIHDOANAVQTTDRIDLEGAKGGLFVRPVAVDL--TYIPVVG-----HA 275
DB 230 AQVRVAAAAYETAVGLTVPPVPI--AENRAELMILATNLGQNTPAIAVNEAEYGEWMA 287

QY 276 LSAAFQAPFCAGAVVGGALVYLVKTLINATQLKLAKLAELVAAAIAD-IISDVAD 334
DB 288 QDAAMFQYAAATATATATALLPEEPAPMTSAGLLLEQAAVEEASDTAAANQOLMNNVPQ 347

QY 335 IIRGI-----LGEVWEFITNALNKLKELWDLTGWTGLFSRGSNLSFFPAG 382
DB 348 ALQOLAQPTGTPSSKLGGLWKTVPSPHRSPI-----SNMYSMANN 388

QY 383 VPGLTGNTSLGSLQVTLFGAAGLSASGLAHADSLASSASLPALAGIG---CGSGFGG--- 437
DB 389 HMGNT--NSGVSTNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGGV 446

QY 438 -----LPSLAQVHA--ASTQALRPADG-----PVGAAABQVGGQSLVSAQSQCM 483
DB 447 AANLGRAASVCSLSVPOAAANQAVTPAARALPLTSLISAERGFGQM-----L 496

QY 484 GG-FVGMGGHPSPSGAS 499
DB 497 GGLPVGMGARAGGGLS 513

RESULT 11
US-09-287-849-28
; Sequence 28, Application US/09287849
; Patent No. US20020009459A1

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040~
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 28
LENGTH: 231
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-28

Query Match 27.0%; Score 739; DB 9; Length 231;
Best Local Similarity 100.0%; Pred. No. 8.1e-50;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 60
DB 2 MHHHHHTAASDNFQSQGGGFAIPGQAMATAGQIRSGGSPVHIGPTAFGLGVVD 61

QY 61 NNGNGARVQRVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 62 NNGNGARVQRVGSAAPASLGISTGVTAVDGPINSATAMADALNGHHPGDVSVTWQ 121

QY 121 TKSGGTRTGNVTLAEGPPAEF 141
DB 122 TKSGGTRTGNVTLAEGPPAEF 142

RESULT 12
US-10-359-460-28
; Sequence 28, Application US/10359460
; Publication No. US20030147911A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-28

Query Match
Best Local Similarity 27.0%; Score 739; DB 12; Length 231;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 60
DB 2 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 61
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
DB 62 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 121
QY 121 TKSOGTGTGNVTLAEGPPAEF 141
DB 122 TKSOGTGTGNVTLAEGPPAEF 142

RESULT 13
US-09-841-132-337
; Sequence 337, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-337

Query Match
Best Local Similarity 26.9%; Score 737.5; DB 9; Length 585;
Matches 231; Conservative 40; Mismatches 174; Indels 151; Gaps 22;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 56
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
DB 57 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 116
QY 121 TKSOGTGTGNVTLAEGPPAEF
DB 117 TKSOGTGTGNVTLAEGPPAEF
QY 174 NOGQILYSSLE---YFEKA-----VPRGMSRAFIIDPTISAIDGLYDLGIG-IP 173
DB 117 TKSOGTGTGNVTLAEGPPAEFCRYPHWRPLDQVSESTPSPD-----DVLGKG 168
QY 174 NOGQILYSSLE---YFEKA-----LEELAAAPFGDGLGSAADKYAG-----KRNHVNFFQ 222
DB 117 TKSOGTGTGNVTLAEGPPAEFCRYPHWRPLDQVSESTPSPD-----DVLGKG 168
QY 174 NOGQILYSSLE---YFEKA-----TRDIL--EGAKKG-----LEFVR 260
DB 174 VEGGINFQDLLEIRIKYKAGTETKKTILPSLQAQASANADAWASSPQSGS----- 228
QY 218 VNFQELADLRQLISLIHQANAVQTRDILEGAKKGLFVRPVA-----DLTY 268

DB 229 -----ATTVSDSGSDSSGSDTSETVPVTAKGGLYTDKNLSI 267
QY 269 IPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLINATQLIKLAKLAELVAAAIADI 328
DB 268 TNITG-IIEIANKKATDVGGAYVKT-----TCENSHR-LQFLKNSDCKGG----- 314
QY 329 ISDVADIIKIGELVWEFITNALNGKELWDLKTGWVT--GLFSRG-----WSNLESF- 379
DB 315 -----GIYGEDNITLSN-LTG-KTLFQENTAKEEGGLFIKGTDKALWTGLDSFC 363
QY 380 -----FAGVPLGTGATSLSQVTG---LFGAAGLSASSG 410
DB 364 LIINTSEKGGGAFVTKIEISQTVTSVETIPGIT-PVHGETVITGNKSTGGGGVCTKR 422
QY 411 LAHADSLASASLPALAGIGGGGGLPGLPSLAQVHAASRQALRPRADGPVGAARQVGG 470
DB 423 LALSNLQISISGNSAAENGG-----AHTCPDSFPTAD-----TAQPPAA 463
QY 471 QSOLVSAQSQSGMGPGVGMGHPSPSSGASKGTTTKYSEGAAGTDEARAPVEAD 526
DB 464 ASAASTPKS-----APVSTALSTPSSSTVSSLTLLAASSQASPATSKETQDPNAD 515

RESULT 14
US-09-841-132-353
; Sequence 353, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 353
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-353

Query Match
Best Local Similarity 26.6%; Score 728; DB 9; Length 583;
Matches 228; Conservative 49; Mismatches 186; Indels 140; Gaps 21;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPTVHIGPTAFGLGLGVD 56
QY 61 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 120
DB 57 NNGNGARVQVVGSAPASLSIGTGDVITAVDGPINSATAMADALNGHHHPGDVISVTWQ 116
QY 121 TKSOGTGTGNVTLAEGPPAEF-----LVPRGMSRAFIIDPTISAIDGLYDLGIG 171
DB 117 TKSOGTGTGNVTLAEGPPAEFCRYPHWRPLDQVSESTPSPD-----DVLGKG 168
QY 172 IPNOGQILYSSLEYPEKALEELAAAPFGDGLGSAADKYAG-----KRNHVNFFQ 222
DB 169 -----GGI-----YTEKSLTITGITIDFVSNITADSGAGVFTKENLSCTNLSLQFLK 218
QY 223 ELADLD-----RQLISLIHQANAVQTRDIL--EGAKKG-----LEFVR 260
DB 219 NSAGOHGGGAVVTQMSVNTTSSITPPVLVGEVIFSENTAKHGCGGICTNKLKSLNKL 278
QY 261 PVAV-----DLTYIPVVGHALSAAFOAPFCAGAMAVVGGALAYLVVKTLIN 306
DB 279 TVTLTKSAKESGGAIPTDLASIPITDTPESSTSSSPASTPPEVVASA----- 327

QY 307 ATQLKLLAKLAELVAADIIISVDAD-----IIKGIIEVWEFITNALNG 353
Db 328 --KINRFFASTAEPAPSLTEAESDQTQTETSDTNSDIDVSIENILNVAINTQNTSAKGG 385
QY 354 LKELWDKLTGWVTGLFSR--GHSNLESFFAGVPGITGATS-----GLSQVTGL-FGAAGL 405
Db 386 -----GAIYKKAKLSRINNLE-----LSGNSSQDVGGGLCLTESVEFDAIG- 427
QY 406 SASSGLAHADSLASSASL-----PALAGIGGSGFGGLPSLAQVHAASLROALRPADGP 460
Db 428 ---SLLSHVNSAAKEGGVTHSKTVLSNLKFTFPADNTVKAIVESTPEAPEEPPVEGE 484
QY 461 VGAAP-----QVGGQSOLVSAQSGQMGPGVGMGWHFSSGASKGTTTTKKYSEGAAGTE 516
Db 485 ESTATENPNSNTEGSSANTNLESGQDTADTGTGVVNNESQDTSDTGNAESGEQLQDSTQ 544
QY 517 DAE 519
Db 545 SNE 547

RESULT 15
US-09-841-132-333
; Sequence 333, Application US/09841132
; Patent No. US20020061648A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 333
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-333

Query Match 26.4%; Score 722.5; DB 9; Length 518;
Best Local Similarity 37.3%; Pred. No. 4.8e-48;
Matches 209; Conservative 44; Mismatches 132; Indels 175; Gaps 21;
QY 1 MHHHHHTAASDNFQLSQGGQFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGIVVD 60
Db 1 MHHHHHTAASDNFQLSQGGQFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGIVVD 56
QY 61 NNGNGARVQVGSAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVTWQ 120
Db 57 NNGNGARVQVGSAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVTWQ 116
QY 121 TKSQGTGTGNTVLAEGPPAEF-----LVPRGMSRAP-IIDPTISAIIDGLYDLIGI 170
Db 117 TKSQGTGTGNTVLAEGPPAEFCRYPYSHWRPLMKWLSATAVFAAVLPSVS-----GF 167
QY 171 GIPNOGILYSSLE-----YFEKALEELAAFPDGMWG-----SAADKYAGKRNHV 218
Db 168 CFPEPEKELNFSRVETSSSTFTTIGEAGAYIVSNGASFKEFTNPTTDTTTPINSNS 227
QY 219 NFFQELADLRQLISLHDOANAVQTRDILEGAKGLEFVRVAVDLTYIPVVGHALSA 278
Db 228 SSSGETAS-----VSEDSDSSTTTPD-----PKG-----251
QY 279 AFQAPFCAGAMVVGGLAYLVVVKTLINATQLLKLAKLAELVAAAIADIISVDADI 338
Db 252 -----GGA-----FYNH-----SG 261
QY 339 ILGEVWEFITNALMGLKELMDKLTGWVTGLFSRGSNLESFFAGVPLGTGATSLQVGTG 398
Db 262 VLSFMTSRSGTEGSLTISEI--KMTGEGGAIFSQ-----ELLFTDLTSLT-IQNNLSQLSG 314

QY 399 --LFGAAGLSASSGLAHADSLASSASL-----PALAGIGGSGFGGLPSLAQVHAASLROALRP 456
Db 315 GAIFGGSTISL-SGITKATFSCNSAEP-----APVKRTPBK 351
QY 457 ADGPVGAALAEQVGGQSOLVSAQSGQMGPGVGMGWHFSSGASKGTTTTKKYSE-----G 510
Db 352 AQ-----TASETSGSS-----SSGNDVSVS-----PSSSRAEPAAANLQSHFICATAT 395
QY 511 AAAGTEDAERAPVEADAGGG 530
Db 396 PAAQTDITETSTPSHKPGSGG 415

Search completed: November 21, 2003, 16:38:19
Job time : 22.0222 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:57:31 ; Search time 11.2444 Seconds
(without alignments)
4609.825 Million cell updates/sec

Title: US-09-688-672A-64

Perfect score: 2737

Sequence: 1 MHHHHTAASDNFQLSGG.....RAPVEADAGGQKVLVRNV 539

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1963	71.7	392	2 A70957	hypothetical prote
2	1263	46.1	394	2 S72814	hypothetical prote
3	676	24.7	355	2 F70993	probable serine pr
4	474.5	17.3	361	2 S47170	hypothetical prote
5	460.5	16.8	354	2 A87242	probable secreted
6	385.5	14.1	402	2 E70656	hypothetical prote
7	199.5	7.3	464	2 C70821	probable serine pr
8	188.5	6.9	382	2 H86930	probable serine pr
9	188.5	6.9	452	2 T45448	probable serine pr
10	173	6.3	837	2 H72802	minor tail subunit
11	169	6.2	496	2 H70839	hypothetical glyci
12	168	6.1	1417	2 A83080	hypothetical glyci
13	167	6.1	916	2 T03323	gene 116 protein -
14	166.5	6.1	1018	2 H83135	probable adhesin p
15	165	6.0	864	1 EART	elastin precursor
16	164	6.0	837	2 S30971	gene 26 protein -
17	159.5	5.8	860	1 FAMS	elastin precursor
18	159	5.8	1467	2 A75564	conserved hypothet
19	154	5.6	1329	2 E70917	hypothetical glyci
20	152.5	5.6	914	2 H70987	hypothetical glyci
21	152	5.6	409	2 A70647	probable PPE prote
22	151.5	5.5	940	2 F84089	phage-related prot
23	151	5.5	588	2 F70971	hypothetical glyci
24	150	5.5	439	2 D70954	hypothetical glyci
25	150	5.5	741	2 G70917	hypothetical glyci
26	150	5.5	801	2 F70824	hypothetical glyci
27	150	5.5	1306	2 A70934	hypothetical glyci
28	150	5.5	1517	2 T13329	hypothetical prote
29	149.5	5.5	749	2 A70812	hypothetical glyci

RESULT 1

A70957

hypothetical protein RV3616c - Mycobacterium tuberculosis (strain H37Rv)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: A70957

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R/Connor, R.; Davies, R.; Devlin, K.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

R/Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: A70957

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-392 <COL>

A/Cross-references: GB:Z95557; GB:AL123456; NID:G3242276; PIDN:CAB08950.1; PID:e316833; I

A/Experimental source: strain H37Rv

C/Genetics:

A/Gene: RV3616c

Query Match 71.7%; Score 1963; DB 2; Length 392;

Best Local Similarity 99.7%; Pred. No. 2.4e-108;

Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 207

Db 1 MSRAFIIDPTTISAIDGLYDLGIGIPNOGGILYSSLEYFEKALELAAAFPGDGLGSA 60

QY 208 DYAGKRNHNVPFQELADRLQLSLIHQANAVQTTTRDILEGAKKGLFVRPVDLT 267

Db 61 DYAGKRNHNVPFQELADRLQLSLIHQANAVQTTTRDILEGAKKGLFVRPVDLT 120

QY 268 YPVPVGHLSAAFAFPFCAGAMAVGGALVIVKTLINATQLLAKLAELVAAATAD 327

Db 121 YPVPVGHLSAAFAFPFCAGAMAVGGALVIVKTLINATQLLAKLAELVAAATAD 180

QY 328 IISDVADIKIGLGEVWEFFITNALNGLKELDKLTGWVTGLFSRGSNLSFFAGVPGIT 387

Db 181 IISDVADIKIGLGEVWEFFITNALNGLKELDKLTGWVTGLFSRGSNLSFFAGVPGIT 240

QY 388 GATSGLSQVTGLFGAGLSASSGLAHADSLASSALPALAGIGGSGFGGIPSLAQVHAA 447

Db 241 GATSGLSQVTGLFGAGLSASSGLAHADSLASSALPALAGIGGSGFGGIPSLAQVHAA 300

QY 448 STRQALRPADGPVGAAGAEQVGGSQLVSAQSGQMGPGVGMGWHPRSSGASKGTTTKY 507

Db 301 STRQALRPADGPVGAAGAEQVGGSQLVSAQSGQMGPGVGMGWHPRSSGASKGTTTKY 360

QY 508 SEGAACTEDERAPVEADAGGQKVLVRNV 539

|||||

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2

S72814

hypothetical protein B1620.C2.213 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72814

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1620.

A:Reference number: S72584

A:Accession: S72814

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <SMI>

A:Cross-references: EMBL:U00015; NID:g466931; PIDN:AAC43223.1; PID:g466935

Query Match 46.1%; Score 1263; DB 2; Length 394;

Best Local Similarity 62.9%; Pred. No. 3.6e-67;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTIGSIDGLYDLGIGIPNQGILYSSLEYFEKALELEAAAFPGDGLGSA 207

DB 1 MSQAFIIDPTLKAIEAWHALLGIGVNDGGVLYSSLSFEKALEHAAAFPGDGLGSA 60

QY 208 DKYAGQNRHNVFQELADLDQLISLHDQANAVOTTDILEGAKKGLFVRPVAVDLT 267

DB 61 DKYAGQNRKRVDFQELADLDKELIELIHQANSVOTTRGILDGAKKALLFVRPVAVDLN 120

QY 268 YIPVWGHLSAAPQAPFCAGAMVVGALAYLVVKTLLINATQILKLAKLAEVAAAIAD 327

DB 121 YIPVGSVNSASIQACAAAMAASVGLAYLLVQTAIHTAKFVALLARLALLASAVAD 180

QY 328 IISDVADIINGLGEVWEFITNALNGKLWDLKLTGWVTLFSGWNSNLESFFAGVPGLT 387

DB 181 VSDGVAIIKGIVDHLWHFETAGALTGLKDIKVEIHHWFGLFSHWSRLHSFPGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLASSGLAHADSLASSLPALAGIGGSGFGLPSLAQVHAA 447

DB 241 GATSGLSQVTLGFGVGLAGSSGLLSSTENLPSLAGVAGLGLSLQLAQHLAA 300

QY 448 STRQALRPADGPGVGAAGVQGSQSLVSAQSGQMGVPGVMGHPSSGASK--GTTTK 505

DB 301 STRQTRSQAGVSAELSTEQFGQEPVSAQSGQMGSGMGGMTPASTKSKDKERKK 360

QY 506 KYSEGAAGCTDAERAPVEADAGGQKVLVRNV 539

DB 361 KYSEGAAGTDAERAPIEVQSGGKRAUQHVV 394

RESULT 3

F70383

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70383

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70383

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:296071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 24.7%; Score 676; DB 2; Length 355;

Best Local Similarity 99.2%; Pred. No. 1e-32; Indels 0; Gaps 0;

Matches 132; Conservative 1; Mismatches 0;

QY 7 HTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGA 66

DB 223 NTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGA 282

QY 67 RVQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 126

DB 283 RVQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 342

QY 127 RTGNVTLAEGPPA 139

DB 343 RTGNVTLAEGPPA 355

RESULT 4

S47170

hypothetical protein 34K - Mycobacterium paratuberculosis

C:Species: Mycobacterium paratuberculosis

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002

C:Accession: S47170

R:Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.

submitted to the EMBL Data Library, June 1993

A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratub

A:Reference number: S47170

A:Accession: S47170

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <CAM>

A:Cross-references: EMBL:Z23092; NID:g505550; PIDN:CAA80638.1; PID:g505551

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 17.3%; Score 474.5; DB 2; Length 361;

Best Local Similarity 69.7%; Pred. No. 7.6e-21;

Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

QY 8 TAAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGAR 67

DB 231 TAATDSYKMS--GGQGFAPIGRAMAVANQIRSGAGSNTVHIGPTAFGLGLGVVNDNNGAR 289

QY 68 VQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 127

DB 290 VQVVNTGPAAGIAPGDVITVTPINGATSMTEVLVPHHPGDTIAVHFRSVDGGER 349

QY 128 TGNVTLAEGPPA 139

DB 350 TANITLAEAGPPA 361

RESULT 5

A87242

probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C:Accession: A87242

R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, R.; Davies, R.M.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: A87242

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-354 <STO>

A:Cross-references: GB:AL450380; NID:gl3093863; PIDN:CAC32191.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML2659

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2

S72814

hypothetical protein B1620.C2.213 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72814

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1620.

A:Reference number: S72584

A:Accession: S72814

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <SMI>

A:Cross-references: EMBL:U00015; NID:g466931; PIDN:AAC43223.1; PID:g466935

Query Match 46.1%; Score 1263; DB 2; Length 394;

Best Local Similarity 62.9%; Pred. No. 3.6e-67;

Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTIGSIDGLYDLGIGIPNQGILYSSLEYFEKALELEAAAFPGDGLGSA 207

DB 1 MSQAFIIDPTLKAIEAWHALLGIGVNDGGVLYSSLSFEKALEHAAAFPGDGLGSA 60

QY 208 DKYAGQNRHNVFQELADLDQLISLHDQANAVOTTDILEGAKKGLFVRPVAVDLT 267

DB 61 DKYAGQNRKRVDFQELADLDKELIELIHQANSVOTTRGILDGAKKALLFVRPVAVDLN 120

QY 268 YIPVWGHLSAAPQAPFCAGAMVVGALAYLVVKTLLINATQILKLAKLAEVAAAIAD 327

DB 121 YIPVGSVNSASIQACAAAMAASVGLAYLLVQTAIHTAKFVALLARLALLASAVAD 180

QY 328 IISDVADIINGLGEVWEFITNALNGKLWDLKLTGWVTLFSGWNSNLESFFAGVPGLT 387

DB 181 VSDGVAIIKGIVDHLWHFETAGALTGLKDIKVEIHHWFGLFSHWSRLHSFPGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLASSGLAHADSLASSLPALAGIGGSGFGLPSLAQVHAA 447

DB 241 GATSGLSQVTLGFGVGLAGSSGLLSSTENLPSLAGVAGLGLSLQLAQHLAA 300

QY 448 STRQALRPADGPGVGAAGVQGSQSLVSAQSGQMGVPGVMGHPSSGASK--GTTTK 505

DB 301 STRQTRSQAGVSAELSTEQFGQEPVSAQSGQMGSGMGGMTPASTKSKDKERKK 360

QY 506 KYSEGAAGCTDAERAPVEADAGGQKVLVRNV 539

DB 361 KYSEGAAGTDAERAPIEVQSGGKRAUQHVV 394

RESULT 3

F70383

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70383

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.B.; Taylor, K.; Whitehead, S.; Barrall, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70383

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <COL>

A:Cross-references: GB:296071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: pepA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 24.7%; Score 676; DB 2; Length 355;

Best Local Similarity 99.2%; Pred. No. 1e-32; Indels 0; Gaps 0;

Matches 132; Conservative 1; Mismatches 0;

QY 7 HTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGA 66

DB 223 NTAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGA 282

QY 67 RVQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 126

DB 283 RVQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 342

QY 127 RTGNVTLAEGPPA 139

DB 343 RTGNVTLAEGPPA 355

RESULT 4

S47170

hypothetical protein 34K - Mycobacterium paratuberculosis

C:Species: Mycobacterium paratuberculosis

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002

C:Accession: S47170

R:Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.

submitted to the EMBL Data Library, June 1993

A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratub

A:Reference number: S47170

A:Accession: S47170

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <CAM>

A:Cross-references: EMBL:Z23092; NID:g505550; PIDN:CAA80638.1; PID:g505551

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

Query Match 17.3%; Score 474.5; DB 2; Length 361;

Best Local Similarity 69.7%; Pred. No. 7.6e-21;

Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

QY 8 TAAASDNFQSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVNDNNGAR 67

DB 231 TAATDSYKMS--GGQGFAPIGRAMAVANQIRSGAGSNTVHIGPTAFGLGLGVVNDNNGAR 289

QY 68 VQVVGSAPAAASLGISTGDVITAVDGAIPNSATADALNGHHHPGDVLSVTWTKSGGT 127

DB 290 VQVVNTGPAAGIAPGDVITVTPINGATSMTEVLVPHHPGDTIAVHFRSVDGGER 349

QY 128 TGNVTLAEGPPA 139

DB 350 TANITLAEAGPPA 361

RESULT 5

A87242

probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C:Accession: A87242

R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, R.; Davies, R.M.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: A87242

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-354 <STO>

A:Cross-references: GB:AL450380; NID:gl3093863; PIDN:CAC32191.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML2659

Best Local Similarity 39.8%; Pred. No. 0.00054; Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 9 AASDNFQLSGGGGFAIFIGOAMAIAGQRISGGGSPVHIGFTAFGLGVVDNNNGN-GAR 67
| | : | | ||||| : | | : | | : | | :
| | : | | ||||| : | | : | | : | | :

QY	262	VAVELTVIPVGHALS---	AAFOAPFCAGAMAVVG----	293
		:::::	:::::	
Dd	11	VAAAATHLAGIGSALSTANAAAAAP--TTALSVAGADEVSVLTAALFEAYAQEYQALSAQ	68	
		:::::	:::::	
QY	294	-----GALAYLVVKLINATQLLKLAELVAAAIADIISDVADII---	336	
		:::::	:::::	
Dd	69	ALATHDOFVQALNMGVCYAAAE-ANAATPL-----QALQTVOQNVLTVVNAP	115	
		:::::	:::::	
QY	337	-KGILGEVMEFITNALNGKLKELDKLTGWVTG-----LFSRGWSNLSEFFAGVPGLTGATS	391	
		:::::	:::::	
Dd	116	TQALLGR--PIINGANGLPN-----TGDGGPGGLLFNGNGN-----GGSG	155	
		:::::	:::::	
QY	392	GLSQVTGLFGNAAGL--SASSGLAHADSASALPA-----LAGIGGSFGGLPSLAQV	444	
		:::::	:::::	
Dd	156	GVDQAGNGGAAGLIGNGGSGVGPGIAGSAGGAGGAGGLIFGNGPFGGAGGITGTG-	214	
		:::::	:::::	
QY	445	HAASTRQALRPDADPGVAAAEOVGGOSQLVSAGSOGSGMGPVMGMGMHPSSCASKGTTT	504	
		:::::	:::::	
Dd	215	-----GGPGGAGGNAIG-----LFGSGGTGGMGVGMGVGNAGNGGTA	257	
		:::::	:::::	
QY	505	KKYSEGAAAGTEDAERPVRADAGG	530	
		:::::	:::::	
Dd	258	GLFHGGHAGGAGGIGCSADGSLGGGGG	283	
		:::::	:::::	

RESULT 12

A83080
hypothetical protein PA4541 [imported] - Pseudomonas aeruginosa (strain PAO1)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83080
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A/Reference number: A82950; UID:20437337; PMID:10984043
A/Accession: A83080
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1417 <STO>
A/Cross-references: GB:AEO04867; GB:AEO04091; NID:g9950769; PIDN:AAG07929.1; GSPDB:GN001
A/Experimental source: strain PAO1
C/Genetics:
A/Gene: PA4541

Query Match	6.1%; Score 168; DB 2; Length 1417;
Best Local Similarity	22.1%; Pred. No. 0.042;
Matches	139; Conservative 77; Mismatches 206; Indels 206; Gaps 31;
QY	19 GGQFAIPICGMAAIAG-----QTRSG 41
Db	782 GGNSSFLISNAGAFGGTFDGLGNTINDLAVYGTGAYGGLFSVNRGTLRNLNERISADG 841
QY	42 GSPT---VHTGPTAFLGLGVDDN-NGNGARVQRVWGSAPAASLGISTGVDVITAVDGAPIN 97
Db	842 AQATHYVNVQVGSAAVNLGRIDNVNASDIRI-----AAASKNLSGG-LVALNLGSD 893
QY	98 SATAMADALNGHPGDVIVTWQKSGITGNTVLAEG-----PPAEFLVP----- 144
Db	894 NASASGTLVGNRH-----TYAL--GGLAAENISTARGVASISNRRADFALSGQLKDH 944
QY	145 -----RGMSRAFIIDPTISAIDGLYDLLGIGIPNQGIL-YSSLEYFEKALE 191
Db	945 SHYGAGGLVGRNRRGLIRS-----SGSQCTLSLSGHGM-NLGLLVGYSS-----AGGLA 992
QY	192 ELAAA--FPDGMGLSAADKYAGKRNHNHVFQELADLRQLLSLTHDCAV--QTTRD 247
Db	993 DVASVDVSGNQEG---LYGGLIGLVN-----SGIAHATASGKVRGTDAE 1036
QY	248 ILGSAKGLFEVRPVAVDLTYIPVWGHALSAAPFCAGAMAV-----VGGALAY----- 298

```

Db      1037  ALGG-----LIGRNUNNAINNAHGDVSLQAGRYLGGDIGHNQAG 1077
Qy      299  -----LVVKTILNATQLKLAKLAELVAAAIAIISDVADIKGL----- 340
Db      1078  NLNVSTSGNLSGGSLIQAGGLIGLNANASLVNASAKGNVATRCAEAVGGLLGENLYGSV 1137
Qy      341  -----GEWWEFITNALNGKLKELWDKLTGWVTGLFSRGSNLSFFAGVPGLTGATSGLS 394
Db      1138  INGSASGEVTDGSGKTGLGL--IGSNLGNHNLKASGMVN-----AGAN---S 1181
Qy      395  QVTGLFGAAGLSASSGLAHADSLASSASLPALAG--IGGSGFGGLPLSLAQVHAASTROA 452
Db      1182  DVGGLLG-----HNREGNHS--TLAASGNVTGKGSRVGGLVGYNDAAASLTNVASGNVSA 1235
Qy      453  LRPRADGVPGAABOVGQSQSLVSAQG-----SQMGGSPVGMG-----GMHPSSGASKG 501
Db      1236  SGRSAIG--GLIGSLRGSLMLASSHGIVYNDKTSHLGLVGRGENTSIIRSAKASGAVSG 1293
Qy      502  TTTKKYSEGAAGTDAERAPVEADAGG 529
Db      1294  GAGLR--AGGLVSLGEGWQALILGASAGG 1320

RESULT 13
T03323
gene l16 protein - Lactococcus phage b11170
C:Species: Lactococcus phage b11170
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 05-Nov-1999
C:Accession: T03323
R:Crutz-Le Coq, A.M.; Cesselin, B.; Commissaire, J.; Anba, J.; Kyriakidis, S.; Chopin, M.
submitted to the EMBL Data Library, June 1997
A:Description: Sequence and organization of the lactococcal isometric b11170 phage genome
A:Reference number: Z14903
A:Accession: T03323
A:Status: Preliminary; translated from GB/EMBL/DBEJ
A:Molecule type: DNA
A:Residues: 1-916 <CRU>
A:Cross-references: EMBL:AF009630; NID:G3282260; PIDN:AAC27195.1; PID:G3282276
C:Genetics:
A:Gene: l16

```

```

Query Match      6.1%; Score 167; DB 2; Length 916;
Best Local Similarity 21.8%; Pred. No. 0.029;
Matches 119; Conservative 70; Mismatches 189; Indels 168; Gaps 25;

QY      51  TAFELGLG-----VVDNN---CNGARVQEVV---GSAPAAISGISGDVITAV 91
Db      229  TTFIGLGDSAKTAVSKTEALVKANQAFGTGNNLKGWQVQYGMSAAG-----KV 278
QY      92  DGAPINSATAMADALNGHHPGDVI SVTWQTKSGGT-----RTGNVTL-----AEG 136
Db      279  TAENINQLTDNNTALGASLQKDTVMQMNPLKQYGSFNEAVSAGAVSM DMLDKAMQRAADG 338
QY      137  PPASEFLVRGSGS-----MSRAFIIDPTISA---IDGLYDILGIGIPNQGGILYSLEY 185
Db      339  SSSATKTIRDTWSGFNEDLSQALI--PTLEALTVPVINALIDKMDMDGK GAGKAIENVVKY 396
QY      186  FE---KALEE-----LAAAPFG-DGWLGSAAKDVAG---KN---RNHV 218
Db      397  FQDLFKQLQQNGAITQPSAIWNDLKSAFGSVIGIIGNLKSAFVDES TSKNSTSVENVA 456
QY      219  NFFQELADLRQLSLIHDQNAVOTTRDILEGAKKGLEFVRPVAVDL-----T 267
Db      457  NTISSLANKFADITTKKADTFGKISEKSA MD AIKVALVALAGAFVAMKVINGIIRKAYET 516
QY      268  YIPVW--GHALSAAFQAPFCAGAMAVVGALAYLVVKTILINATQL--LKLAKLAELVA 322
Db      517  YNKIVEAGTTIQGAFNALMAVNPVLLGIGIALAAVAGLVYFQTGTETG KANWASFVDLFLK 576
QY      323  AA-----TADISDVADIKGI-----LGEVWEFI-----347
Db      577  SMDGIVSFSGIGQWFADINWGAVDGAKTQWGLVDWFIQVQGIQNTWNGITTFPTTL 636

```

QY 348 -TNALNGLKELDKLGTWVTLGLF-----SRGWSNLESPFAGVPGITG 388
Db 637 WTVVIGIGSVGVGTGFGIFDAVSVVTVFSAIGGFASSAMNVLVSWSAVAGPFG 696
QY 389 ----ATSGLSQVTLGLFGAAGLASSGLAHADSLASSAS-----LPALAGIGGG--SGFG 436
Db 697 GIFNAVSGV--VSSVFAISGPFASSAMGVVGVVINSVSGFGFINSVSSVGVVFSALG 754
QY 437 GLPSLA 442
Db 755 GFASNA 760
RESULT 14
H83135
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83135
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1018 <STO>
A:Cross-references: GB:AE004824; GB:AE004091; NID:g9950277; PIDN:AAG07469.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4082
Query Match 6.1%; Score 166.5; DB 2; Length 1018;
Best Local Similarity 23.7%; Pred. No. 0.035;
Matches 149; Conservative 75; Mismatches 259; Indels 145; Gaps 28;
QY 5 HHTAASDNFQLSQGGGFAIPICQAMATAGQIRSGGSPVTHIGTAPLGLGVVDN--- 61
Db 174 YRFTGTPSTNGVLNHHGGAITAAEGGSIALGCAQVDNRG---TV-LAQMGGVGLGAGSDTL 229
QY 62 NNGARVORVVGAPAAUSLGISTGDIVTAVDGAPINSATAMADALNG---HHPGDVISVT 118
Db 230 NFGNKLDIRVDAGVANALASNGLLKA-DGRVLMARTANALLNTVNSQGA---- 284
QY 119 WQKSGGTRTGNVTLAEGPAPFLVPRGMSRAFIIDPTISADGLYLLGIGIPNQGI 178
Db 285 -EARSLRGKNGRIVLDGGPGKVMV--GGALS-----ANALNG-----FHGGGT 325
QY 179 LYSLEVFKEALEELAAAPFGDGLGS---AADKY-----AGKRNHNVFFOBLADL 227
Db 326 VEVRGQAVEALGTQNTLASNGLWTWKAADKIDVRSAVSDGTVTVADTLRNLAST 385
QY 228 DRQLISLIHD-----QANAVQTRIDILEGAKGLFEVRPVAVDLTY 268
Db 386 NIELVSTKGBDLDDGSVNWASGNRLGLGSAADTLNGLNASAKAGLEKAEGAIDIND 445
QY 269 IPVVGHALSA-----AFQAPFCAGAMAV-VGGALAVLVYKTLINATQLLKLA 315
Db 446 KIVLGGAGSALAMDAGSGHRVNGTASVSLAGANATVYSGGYVTVQNLQAQAINKND 505
QY 316 KLAELVAAAIADIISDVADIKIGILGEVWEFFITNALNGLKELWDKL-----TGWVTGLFGR 371
Db 506 GLYVLGNILGG--SYCYCTALQSIGGPAGVF-SGTLDDLGNLSIGNLSINTGNVGLFAR 562
QY 372 G-----WSNL-----BSFPAGVPLGTGATSLGSLQVTLFGAAGLASSGLAHADSLASS 420
Db 563 SSGTSLNKLNNLRVSDNTYSGSPSSIGALVNGS-----GRIANVSAGSVVSGRLRS 617
QY 421 ASLPALAG-----IGGSGFGGLPLSLAQVHAASTROALPRADGPFVGAAGVQVQSOL 474
Db 618 NALGGLVGRNISGOIANASVSGGVGTG-----YAAST-----AVGGLVGENFTTAWGPEAV 667

QY 475 -----VSAQSGQ-----GMGGPVGMG--GMHPSSGASKGTTTKKYSEGAAAG----- 514
Db 668 IENAHSNVHVAQSGSTERNISLGGVGLVGLNAKGMIRASGSGQKVT--YRPLNVGLGLVG 725
QY 515 -----TEDAEAPVPEADAGGQKQVLV 535
Db 726 YNMFHVSVDSSASGQVEAGGAGNTGGLV 753
RESULT 15
EART
elastin precursor - rat
N:Alternate names: tropoelastin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-1991 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C:Accession: A36106; A30878; A36523; S02173; I54172; I68505
R:Pierce, R.A.; Deak, S.B.; Stolle, C.A.; Boyd, C.D.
Biochemistry 29, 9677-9683, 1990
A:Title: Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.
A:Reference number: A36106; MUID:91104868; PMID:1702999
A:Accession: A36106
A:Molecule type: mRNA
A:Residues: 1-864 <PIE>
A:Cross-references: GB:M60647; GB:J05292; NID:G207444; PIDN:AAA42269.1; PID:G207445
R:Deak, S.B.; Pierce, R.A.; Belsky, S.A.; Riley, D.J.; Boyd, C.D.
J. Biol. Chem. 263, 13504-13507, 1988
A:Title: Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.
A:Reference number: A30878; MUID:88330868; PMID:2971041
A:Accession: A30878
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 781-864 <DEA>
A:Cross-references: GB:J04035; NID:G207442; PIDN:AAA42268.1; PID:G207443
R:Franzblau, C.; Pratt, C.A.; Paris, B.; Colaninno, N.M.; Offner, G.D.; Mogayzel Jr., P.
J. Biol. Chem. 264, 15115-15119, 1989
A:Title: Role of tropoelastin fragmentation in elastogenesis in rat smooth muscle cells.
A:Reference number: A36523; MUID:89359327; PMID:2768256
A:Accession: A36523
A:Molecule type: protein
A:Residues: 22-31 <FRA>
R:Rich, C.B.; Foster, J.A.
Arch. Biochem. Biophys. 268, 551-558, 1989
A:Title: Characterization of rat heart tropoelastin.
A:Reference number: S02173; MUID:89117149; PMID:2913947
A:Accession: S02173
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1P, 369-545, 548-764, 770-864 <RIC>
A:Experimental source: heart
R:Pierce, R.A.; Alatawi, A.; Deak, S.B.; Boyd, C.D.
Genomics 12, 651-658, 1992
A:Title: Elements of the rat tropoelastin gene associated with alternative splicing.
A:Reference number: I54172; MUID:92241859; PMID:1572637
A:Accession: I54172
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 264-533 <RES>
A:Cross-references: GB:M86372; NID:G207455; PIDN:AAA42271.1; PID:g554527
A:Accession: I68505
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 558-864 <RE2>
A:Cross-references: GB:M86376; NID:G207459; PIDN:AAA42272.1; PID:G207462
C:Genetics:
A:Introns: 277/1; 292/1; 308/1; 339/1; 359/1; 419/1; 437/1; 467/1; 484/1; 601/1; 621/1; 6
A:Note: the list of introns may be incomplete
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-864/Product: elastin #status predicted <MA>
F:854-859/Disulfide bonds: #status predicted

This Page Blank (uspto)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	173	6.3	836	1	VG26 BPMD2	O64220 mycobacteri
2	165	6.0	864	1	ELS_RAT	Q95372 rattus norv
3	164	6.0	836	1	VG26 BPML5	Q05233 mycobacteri
4	159.5	5.8	860	1	ELS_MOUSE	P54320 mus musculu
5	152.5	5.6	730	1	ELS_HUMAN	P15502 homo sapien
6	152.5	5.6	914	1	WA32 MYCTU	O06794 mycobacteri
7	150	5.5	801	1	Y747 MYCTU	O53810 mycobacteri
8	147	5.4	965	1	DLG1 DROME	P31007 drosophila
9	146.5	5.4	915	1	Y140 MYCTU	Q50594 mycobacteri
10	146	5.3	957	1	Y278 MYCTU	P58877 mycobacteri
11	143.5	5.2	2333	1	PGCA_CANFA	Q28343 canis fami
12	142.5	5.2	2124	1	PGCA RAT	P07897 rattus norv
13	142	5.2	778	1	YQ34 MYCTU	P71933 mycobacteri
14	142	5.2	1901	1	Y208 MYCTU	O53553 mycobacteri
15	141	5.2	2132	1	PGCA_MOUSE	O61282 mus musculu
16	139	5.1	435	1	YU21 MYCTU	O53268 mycobacteri
17	139	5.1	543	1	YP91 MYCTU	O50630 mycobacteri
18	138.5	5.1	419	1	YC58 MYCTU	Q11060 mycobacteri
19	137	5.0	2038	1	FSH DROME	P13709 drosophila
20	136	5.0	747	1	ELS_BOVIN	P04985 bos taurus
21	136	5.0	2415	1	PGCA_HUMAN	P16112 homo sapien
22	135.5	5.0	463	1	YA68 MYCTU	O53416 mycobacteri
23	135.5	5.0	557	1	PAC1_PSESU	Q05053 pseudomonas
24	135.5	5.0	750	1	ELS_CHICK	P07916 gallus gall
25	133.5	4.9	408	1	HTFA_LACLA	Q91a06 lactococcus
26	131.5	4.8	558	1	YJ83 MYCTU	Q10873 mycobacteri
27	130	4.7	498	1	Y118 MYCTU	O50615 mycobacteri
28	130	4.7	3591	1	PHAB_BORPE	P12255 bordetella
29	128.5	4.7	603	1	YD25 MYCTU	Q10637 mycobacteri
30	127.5	4.7	1156	1	GLH4 CAEBL	O76743 caenorhabdi
31	125.5	4.6	557	1	PAC1_PSES3	P15557 pseudomonas
32	125.5	4.6	1758	1	CA24 CAEBL	P17140 caenorhabdi
33	125	4.6	413	1	HTFA_LACHE	Q924h7 lactobacill

QY	227	-----LDRQISLIHDOANAVQTRD-----ILGAKKGLFVFPVAVD	265	CC	Name=2;
Db	356	MPALTSVSSLGNVLGTGLTQAPITALTTPAFTTLADTLGTLTALQALGFLVTVAAE	415	CC	Isoid-Q99372-2; Sequence=VSP_004244;
QY	266	LTVPVVGHALSAFOA-----PFCAGAMAVWGA---	295	CC	Name=3;
Db	416	-----TLGATLITLALQIOMPLTVDSFKQLSEITLVTSGLPYLPQIGFAPQIVGAVIQ	470	CC	Isoid-Q99372-3; Sequence=VSP_004245;
QY	296	LAYLVVKTLLNA-----TOLLKLAKLAELV---AAAIADIIS	330	CC	Name=4;
Db	471	LAFTIISLIPAFQTLIPAIQAQAPLSLQIVQAFKLMFVIVPVVQIVINLAAAVVQAGA	530	CC	Isoid-Q99372-4; Sequence=VSP_004246;
QY	331	DVADIKILGEWEFTNALNGLKELDKLTCWWTGLFSRGSWNLESFAGVGLTGA-	389	CC	Name=5;
Db	531	SIASFLLIGTSRLVGLVADCVGAVAE-----WV-GSWSSGVQVQVSDVFGQLPKIKSW	582	CC	Isoid-Q99372-5; Sequence=VSP_004244, VSP_004245;
QY	390	-----TSGLSQVTGLFGAAGLSAGSLAHADSLASSASLPALAGIGGSGFGGLPS	440	CC	Name=6;
Db	583	FDDAGSLWLEAGKVVQGLINGIGSMISSAVSKAKELASSVK-----NAVTFGLGIHS	635	CC	Isoid-Q99372-6; Sequence=VSP_004245, VSP_004246;
QY	441	LAQVHAASTRQALPRADGPVGAABOVGQSQSVSAQSGQGVGPGVGMGHPGSGASK	500	CC	Name=7;
Db	636	PSRVFABIGQFTAGFGNG-FEEGFQFVIEKAKALAAELSQAM-----BSGVDP-SGILA	688	CC	Isoid-Q99372-7; Sequence=VSP_004244, VSP_004246;
QY	501	GTTTKYSEGAAGTDAERAPVEADA	527	CC	Name=8;
Db	689	GISTKELKQISAALEQERKRIQVEKNA	715	CC	Isoid-Q99372-8; Sequence=VSP_004244, VSP_004245, VSP_004246;
RESULT 2				CC	-!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
ELS_RAT				CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
ID	ELS_RAT	STANDARD;	PRT;	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
AC	Q99372;		864 AA.	CC	the European Bioinformatics Institute. There are no restrictions on its
DT	01-OCT-1996 (Rel. 34, Created)			CC	use by non-profit institutions as long as its content is in no way
DT	01-OCT-1996 (Rel. 34, Last sequence update)			CC	modified and this statement is not removed. Usage by and for commercial
DT	15-SEP-2003 (Rel. 42, Last annotation update)			CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
DE	Elastin precursor (Tropoelastin) (Fragment).			CC	or send an email to license@isb-sib.ch).
GN	ELN.			CC	-----
OS	Rattus norvegicus (Rat).			DR	EMBL; M60647; AAA42269.1; -
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			DR	EMBL; J04035; AAA42268.1; -
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			DR	EMBL; M86372; AAA42271.1; -
OX	NCBI_TaxID=10116;			DR	EMBL; M86355; AAA42271.1; JOINED.
RN	[1]			DR	EMBL; M86363; AAA42271.1; JOINED.
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			DR	EMBL; M86364; AAA42271.1; JOINED.
RX	MEDLINE=91104868; PubMed=1702999;			DR	EMBL; M86366; AAA42271.1; JOINED.
RA	Pierce R.A., Deak S.B., Scollie C.A., Boyd C.D.;			DR	EMBL; M86371; AAA42271.1; JOINED.
RT	"Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";			DR	EMBL; M86376; AAA42272.1; -
RL	Biochemistry 29:19677-19683(1990).			DR	EMBL; M86375; AAA42272.1; JOINED.
RN	[2]			DR	EMBL; M86376; AAA42272.1; JOINED.
RP	SEQUENCE OF 781-864 FROM N.A.			DR	EMBL; M86375; AAA42272.1; JOINED.
RX	MEDLINE=88330868; PubMed=2971041;			DR	PIR; A36106; EART.
RA	Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;			DR	InerPro; IPR003979; tropoelastin.
RT	"Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";			DR	PRINTS; PRO1500; TROPOELASTIN.
RL	J. Biol. Chem. 263:13504-13507(1988).			DR	Structural protein; Connective tissue; Repeat; Signal;
RN	[3]			KW	Alternative splicing.
RP	SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.			FT	NON TER 1 1
RX	MEDLINE=92241859; PubMed=1572637;			FT	SIGNAL <1 21
RA	Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;			FT	CHAIN 22 864
RT	"Elements of the rat tropoelastin gene associated with alternative			FT	DISULFID 854 859
RL	splicing";			FT	VARSPPLIC 263 307
RN	Genomics 12:651-658(1992).			FT	VARSPPLIC 308 308
CC	-!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND			FT	VARSPPLIC 809 823
CC	NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.			FT	SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;
CC	-!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER			FT	Query Match 6.0%; Score 165; DB 1; Length 864;
CC	INTO AN EXTENSIBLE 3D NETWORK.			FT	Best/Local Similarity 23.2%; Pred. No. 0.031;
CC	-!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.			FT	Matches 130; Conservative 48; Mismatches 210; Indels 172; Gaps 26;
CC	-!- ALTERNATIVE PRODUCTS:			QY	19 GQGFAPIGQAMAIAGIIRSGSPVTHIG--PTAFGLGVVDN--NNGGARVQRVVG 74
CC	Event=Alternative splicing; Named isoforms=8;			Db	244 GQQPGVPLGYPIK-APKPGYGLPYTN-GKLPYGVAGAGKAGYPTGTGVGQAATAA 301
CC	Comment=Experimental confirmation may be lacking for some			QY	75 APAAS-LGISGTDVITAVDGAIPINSATAMADALXGHPGDVISTWOTKSGGTRTGNVTL 133
CC	isoforms;			Db	302 AKAKYAGAGGGVLPVGGGGIPGGAGAIPIGIG-----ITGACTPAAAAAK 350
CC	Name=1;			QY	134 AEGPPAEF-----LVPRGMSRA-----FIDPTISAIDGLYLLGI-GIPNOGGILYS 181
CC	Isoid=Q99372-1; Sequence=Displayed;			Db	351 AAACAAYGAAGGLVPGGPGVRVFGAGIPGVGIPGVGGIPGVGGIPGVGGIPGVGG 406
CC				QY	182 SLEYFEKALEELAAAFPGDGLG-----SAADKYAGKNRHNHNVFFELADLRQL 231

```
Db 407 -----PGIGGPIVGGPGAVSPAAAACAAKAAKY----- 436
QY 232 ISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLTYIPVGHLSAFAQAPFCAGAV 291
Db 437 -----GARGVG-----IPTVG--VGAGFFPGYGVGAG 464
QY 292 VGGALAYLVKTLINATOLLKLAELVAALAAIADISDVADIKIGILGEWFEITNAL 351
Db 465 LGGA-----SQAAAAA-----AKAAKYGAGAGTGLGLVPGAV 499
QY 352 NGLKELMDKLTGWVTGLFSRGSNLSFPFAGVPG-LTGATSGLSQVTLFGAAGLSASSG 410
Db 500 PG-----ALPGAVPGALP-----GAVPGALPGAVPGVPGTGGVPGAGTFAAAA 543
QY 411 LAHADSLASSASL---PALAGIGGSGFGGLP---SLAQVHAAT--RQALRPADGPVG 462
Db 544 AAAKAAKAAKQYGLPGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGV 600
QY 463 AAABOVGGQSLVAQGSQ-----GMGGP-VGMGMHPSGA-----SKGITTKKYSE 509
Db 601 GAGTPAAAKSAAKAAKAAQYRAAAGLGAAGVPGVGGVGGVGGVGGVGGVGGVGGV 660
QY 510 GAAAGTEDAERAPVEADAGG 529
Db 661 GAVPGSLAASAAKAAKAAAGG 680

RESULT 3
VG26 BPML5 STANDARD; PRT; 836 AA.
AC Q05233
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Minor tail protein GP26.
GN 26.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RA MEDLINE=93211282; PubMed=8459766;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT a phage system for mycobacterial genetics.";
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18946; CAA79402.1; -
DR PIR; S30971; S30971.0
DR INIT MET 0
SQ SEQUENCE 836 AA; 86258 MW; 52E3040AA42BAD28 CRC64;

Query Match
Best Local Similarity 22.5%; Pred. No. 0.034;
Matches 123; Conservative 75; Mismatches 203; Indels 146; Gaps 24;

QY 85 GDVITA-----VDGAPINSATAMADALNGHHFGDVISVTWQPKSGGTRG-NVTIAEG 136
Db 206 GPMLTATPPLQNVASGLVNNAGSITDVT-QAPGLQIQNLTKTEFFTGLGPVLATG 264
QY 137 PPAEFLVPRGSMRAF--IIDPTISAIDGLDLGIGIPN---QGGI-----LYSSLEY 195
Db 265 TQA-FLTLNSAGANSFGTLAPLOFTNGFNDVMVNTSVNGVFEAGQGLSQTLGSLVNL 323
```

```
QY 186 FEKALBELAAAFPGDGLGSAADKYAGKNRHNVPQEE-----LAD 226
Db 324 FNLMESGLQAM---GQLG-----GPLSTFINGFDLFSVLMPALTSVSGLIGNVLGT 373
QY 227 LDRQLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLTYIPVGHLSAFAQAPFCAGAV 282
Db 374 LGTQLAPIVTALTPAFOTLASTLTGTLTALQALGPI---LTQVATLIGTTLTALQALQ 430
QY 283 -----PFCAGANAVVGA-----LAYLVVKTLINATQLKLL 314
Db 431 PMLPSLMQSFQIQSDVLVTSLAFHPIPALATALQGVAGVQLAPTIIISTLVFA--FVQLV 488
QY 315 AKLAEL-----VAAAIAADIISDVADIKIGILGEWFEI 347
Db 489 PKVAELVPTIVNLVQSOFANLMPVPLAQAALVSVAGAVIQGVSGICGALIGALANLTEL 548
QY 348 TNALNGKELMDKLTGWVTGLFSRGSNLSFPFAGVPG-LTGATSGLSQV 397
Db 549 SNVIK-----KVSEWSS-FSSGAQQAARAAELPGMIQALANLMAIGLQAGKDLVQ 600
QY 398 GLPFGAAGLSASSGLAHADSLASSASL-PALAGIGGSGFGGLPGLSLAQVHAATRAALRPRA 457
Db 601 GLINGIGGVSAAVNRKELASSVA-----GAVKGLGIESPSKLFTEYGGFTAEFG 653
QY 458 DGPVGAARQVGGQSLVAQGSQGMGGPVGVMGMHPSGSGASKGTTTKKYSEGAAGT 517
Db 654 NG-MEAGFKPVIERAKDLAAELSRAM-----ESGTDLP-SGILAGLDQNLKQMLALEE 706
QY 518 AERAPVE 524
Db 707 RKELKVE 713

RESULT 4
ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Lung;
RX MEDLINE=95130089; PubMed=7829080;
RA Wydnar K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.";
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCLEAR LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08210; AAA80155.1; -
DR PIR; A55721; EAMS.
DR MGD; MGI:95317; Eln.
```

```
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL..
FT CHAIN 28 860 ELASTIN.
FT DISULFID 850 855 BY SIMILARITY.
SQ SEQUENCE 860 AA; 71955 MW; 0C0B5AAE1EDD7F1 CRC64;

Query Match 5.8%; Score 159.5; DB 1; Length 860;
Best Local Similarity 23.3%; Pred. No. 0.063;
Matches 118; Conservative 38; Mismatches 186; Indels 165; Gaps 22;

QY 19 GGGGFAIPIGQAWIAAGQIRSGGSTVHIG--PTAFLGLGVVDN--NNGNARVQRVVG 74
DQ 259 GGGQPGVPLGPIK-APKLPGGVGLPYTN-GKLPYGVAGAGKAGYPTGTGVGSGAAAAA 316
QY 75 APAASLGISGDIYITAVDGAIPNSATAMADALNGHPGDVIVTWQTSKGGTRTGNVTIA 134
DQ 317 AKAAKYGAGAGVLPVGGGGIPGGAGATPGIGG-----IAGATPAAAAAAXAAKAA 370
QY 135 EGPPAFBLVPRGMSR---AFIIDPTISAIDGLYDILGICIPNOGILYSSLEYFEKALE 191
DQ 371 KYGAAGLVPGGPGVRLPGAGI--PGVGGIPGVGGIPGVGGPGIGG-----414
QY 192 ELAAATPGDNLG-SAADKYAGKRNHNVFFQELADLDRLQLISLIHQANAVQTTTRDILE 250
DQ 415 --PGIVGGPGAVPAAAKAAKAAKAAKY-----439
QY 251 GAKKGLFVRVPAVDLTYIPVGHLSAAFAQPFACAGAMVVGALAYLWKTINATOL 310
DQ 440 GARGVG-----IPTYG--VGAGGFPGVGVGAGAGLGA-----SPAA 475
QY 311 LKLLAKLAELVAAAIADIISVDADIITKILGEVWEFTITNALNGLKELWKLWGTGTGFS 370
DQ 476 AAAAAAKAAKYGAGG-----AGALGGL-----VFGAVPG---503
QY 371 RGSNLSFPAGVPGLTGATSGLSQVTLGFLGAAGLSAGSLAHADSLASSSL-PALAGI 429
DQ 504 -----ALPGAVPAVPGAGVPG-----AGTPAAAAAAXAAKAAKAGLFGVGVG 548
QY 430 GGGSGFGGLPSLAQVHAASQALRPRADPGVPGAAAEQVGGSQLVSAQSGQMG-GPVG 488
DQ 549 PGGVGVGGIP-----GGVG-----VGGVPGGVGGVGTGIGAGPGG 584
QY 489 MGNM-HPSSGASKGTTTKYSEGAAG 514
DQ 585 LGGAGSPAAAKSAAKAAKAAQYRAAG 611
```

RESULT 5

```
ELS_HUMAN
ID ELS_HUMAN STANDARD; PRT; 730 AA.
AC P15502; Q14233; Q14238;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM B).
RX MEDLINE=87289568; PubMed=3039501;
RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA Rosenbloom J.C., Peltonen L., Rosenbloom J.,
RT "Alternative splicing of human elastin mRNA indicated by sequence
RT analysis of cloned genomic and complementary DNA."
RL Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684 (1987).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Skin fibroblast.
```

RA Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.,
"Cloning of full-length elastin cDNAs from a human skin fibroblast
recombinant cDNA library: further elucidation of alternative splicing
utilizing exon-specific oligonucleotides.";
J. Invest. Dermatol. 91:458-464 (1988).
[3]
SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
TISSUE=Placenta;
MEDLINE=88156138; PubMed=2831431;
Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
Rosenbloom J., Uitto J.,
"Isolation and characterization of human elastin cDNAs, and age-
associated variation in elastin gene expression in cultured skin
fibroblasts.";
Lab. Invest. 58:270-277 (1988).
[4]
SEQUENCE OF 603-730 FROM N.A.
TISSUE=Hippocampus, and Placenta;
MEDLINE=96291399; PubMed=8689688;
Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
Green E.D., Roessel C., Gutowski N.J., Noble M., Atkinson D.L.,
Odelberg S.J., Keating M.T.,
"LIM-kinase hemizygosity implicated in impaired visuospatial
constructive cognition.";
Cell 86:59-69 (1996).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P15502-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P15502-2; Sequence=VSP_004243;
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -!- DISEASE: Haploinsufficiency of ELN may be the cause of certain
cardiovascular and musculo-skeletal abnormalities observed in
Williams-Beuren syndrome (WBS), a rare developmental disorder. It
is a contiguous gene deletion syndrome involving genes from
chromosome band 7q11.23.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M17282; AAC98394.1; JOINED.
EMBL; M16983; AAC98394.1; JOINED.
EMBL; M17265; AAC98394.1; JOINED.
EMBL; M17266; AAC98394.1; JOINED.
EMBL; M17267; AAC98394.1; JOINED.
EMBL; M17268; AAC98394.1; JOINED.
EMBL; M17270; AAC98394.1; JOINED.
EMBL; M17271; AAC98394.1; JOINED.
EMBL; M17272; AAC98394.1; JOINED.
EMBL; M17273; AAC98394.1; JOINED.
EMBL; M17275; AAC98394.1; JOINED.
EMBL; M17276; AAC98394.1; JOINED.
EMBL; M17277; AAC98394.1; JOINED.
EMBL; M17278; AAC98394.1; JOINED.
EMBL; M17279; AAC98394.1; JOINED.
EMBL; M17280; AAC98394.1; JOINED.
EMBL; M17281; AAC98394.1; JOINED.
EMBL; M36860; AAC52382.1; -

```
DR EMBL; M24782; AAB53190.1; -
DR EMBL; U62292; AAB17544.1; -
DR EMBL; X15603; CAA33627.1; -
DR PIR; A32707; EAHU.
DR HSP; P50099; 1ZFU.
DR Genew; HGNC:3327; ELN.
DR MIM; 130160; -.
DR MIM; 194050; -.
DR GO; GO:0005578; C:extracellular matrix; TAS.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0008015; P:circulation; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR GO; GO:0007585; P:respiratory gaseous exchange; TAS.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO1500; TROP0ELASTIN.
KW Structural protein; Connective tissue; Repeat; Signal;
Williams-Beuren syndrome; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 730 ELASTIN.
FT DISULFID 720 725 BY SIMILARITY.
FT VARSPIC 472 477 Missing (in isoform 2).
FT /FTID=VSP_004243.
SQ SEQUENCE 730 AA; 63260 MW; AB06D15BA567AE46 CRC64;
Query Match 5.6%; Score 152.5; DB 1; Length 730;
Best Local Similarity 20.4%; Pred. No. 0.13;
Matches 117; Conservative 46; Mismatches 182; Indels 229; Gaps 23;
QY 16 LSQCGQGAIFIGQMAIAGQIRSGGSPTHIGTAPFLGLGVVDNNG--NGARVQRVVG 73
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
56 LQPGKPLK-PVPGGLAGAGLGLGAPVTF-PGALVPGGVADAAAYAKAKAGAGLG 113
QY 74 SAP-AASGISIGDVIITAVDGPINSATAMADALNGHPGVISVTWTKSGGRTGNVT 132
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 GVPVGLGVAGAVVPO-PGAGVKPGKVGVLPGVPGVLP-----GARFFGVG 164
QY 133 LAEGPAPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGI----PNOGILYSSLEYFEK 188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
165 VLPGVPT-----GAGVKPKAPGVGGAF-----186
QY 189 ALEELAAFPDGGWLGSAADKXVAGKNRNVNFQELADLDRLQLSLTHDQANVQTRDI 248
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 -----AGIPGVGPF-----196
QY 249 LEGAKKGLFVRPVAVDLTYPVVGHALSAAAFQAPFCAGAMAV-----VGGALAYLVVK 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
197 --GPQPG-----VPLGY-PIKAPKLPGGVGLPYTTGKLPYGVPGVAGAKGAGYP 245
QY 303 TLINATQLKLLAKLAELVAIAADIISDVADITKILGEVWEFITNALNGLKELDKLT 362
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 T---GTGVGPQAA--AAAAAKAAAKFGAAGVLPFGVG-----279
QY 363 GWVTGLFRSGMNLSEFFAGVPLGTGATSGLSQVTLGFLGAAGLSASSGLAHADSIASSAS 422
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280 -----AGVPGVPCA-----IPGIGIAVGTPAAAAAANAANAANA 315
QY 423 LPALAG-TGGSGFG-----GLPSL-----AQVHAASTQALPRADGP 460
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
316 YGAAAGLVPGPGFGFGVGVVPGAGVPGVPGAGIPVPGAGIPGAAPVPGVSPFAAAK 375
QY 461 VGAAAEQV-----GQSQLVSAQSGQSGVPGVM 489
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
376 AAKAAKYGARPGVGVGIPTVGAGGFPFGVGVGIPGAGVPGVGVGVPVGVGVPVGV 435
QY 490 GGMHSSGASKGTTTKYSEG--AAAGTEDAPRA 521
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 -GISPEAQAAAAAKAAKYGVGTPAAAAAANAANA 468
RESULT 6
WA22_MYCTU
```

```
ID WA22_MYCTU STANDARD; PRT; 914 AA.
AC O06794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE WAG22 antigen precursor.
GN WAG22 OR RV1759C OR MT1807 OR MTCY28.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean S., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 85.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95890; CAB09322.1; -.
CC EMBL; AE007040; -. NOT_ANNOTATED_CDS.
CC PIR; H70987; H70987.
CC TIGR; MT1807; -.
CC TubercuList; Rv1759c; -.
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC Antigen; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAG22 ANTIGEN.
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8E6AC8 CRC64;
Query Match 5.6%; Score 152.5; DB 1; Length 914;
Best Local Similarity 24.8%; Pred. No. 0.17;
Matches 133; Conservative 40; Mismatches 213; Indels 157; Gaps 25;
QY 19 GGQGFAPIPQAMAIAAGQIRSGGSPTHIGTAPFLGLGVVDNNGARVQRVVGSA 78
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
364 GKGGA--GGAGNAGMLFGSGG-----GGTGGFAGAGGGGAGMLSGSGSG 415
QY 79 SLGISITGDVITAVDGPINSATAMADAL-----NGHHPGDVISTVQTKSGGRTGNVT 133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
416 GAGSGGPGAGTAAGGA---GGAGGAPGLTNGNGNGNGGGE-----SGG--TGGVGG 461
```


SEQUENCE FROM N.A., AND FUNCTION.

RP TISSUE=Embryo;

RC MEDLINE=91330294; PubMed=1651169;

RA Woods D.F., Bryant P.J.;

RT "the disc-large tumor suppressor gene of Drosophila encodes a

RL guanylate kinase homolog localized at septate junctions.";

Cell 66:451-464(1991).

CC -!- FUNCTION: Plays a critical role at septate junctions in cellular

CC growth control during larval development. The presence of a

CC guanylate kinase domain suggests involvement in cellular adhesion

CC as well as signal transduction to control cellular proliferation.

CC Required for maintenance of cell polarity.

CC -!- SUBCELLULAR LOCATION: CYTOSKELETON-ASSOCIATED. LOCATED AT THE

CC CYTOSOLIC FACE OF THE MEMBRANE IN THE CELLULAR BLASTODERM AND

CC BECOMES ASSOCIATED WITH SEPTATE JUNCTIONS WHICH BEGIN TO FORM

CC BETWEEN EPITHELIAL CELLS AT THE TIME OF DORSAL CLOSURE. IN ADULT

CC FLIES, LOCATED AT THE APICAL-LATERAL MEMBRANE BOUNDARY OF

CC EPITHELIAL CELLS.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=1;

CC Comment=A number of isoforms are produced;

CC Name=1;

CC IsoId=FP31007-1; Sequence=Displayed;

CC -!- TISSUE SPECIFICITY: In embryos, expression is seen in epithelial

CC cells and some nervous tissue. In larvae, expression is seen as a

CC belt around salivary glands and imaginal disks, also in

CC proventriculus and parts of the brain. Expressed in adult

CC reproductive tissues.

CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically

CC throughout development.

CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.

CC -!- SIMILARITY: Contains 3 PDZ/DRH domains.

CC -!- SIMILARITY: Contains 1 SH3 domain.

CC -!- SIMILARITY: Contains 1 guanylate kinase-like domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M73529; AAA28468.1; .

DR PIR; A39651; A39651.

DR HSP; P31016; LBFE.

DR FlyBase; FBgn0001624; dgl1.

DR GO; GO:0045179; C:apical cortex; IDA.

DR GO; GO:0016327; C:apicolateral plasma membrane; IDA.

DR GO; GO:0005918; C:septate junction; NAS.

DR GO; GO:0045175; P:basal protein localization; IMP.

DR GO; GO:0007391; P:dorsal closure; NAS.

DR GO; GO:0045197; P:establishment and/or maintenance of epithel. . ; NAS.

DR GO; GO:0016334; P:establishment and/or maintenance of polarit. . ; IGI.

DR GO; GO:0016336; P:establishment and/or maintenance of polarit. . ; NAS.

DR GO; GO:0007399; P:neurogenesis; IMP.

DR GO; GO:0007273; P:regulation of synapse; IMP.

DR InterPro; IPR00619; Guanylate_kin.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00625; Guanylate_kin; 1.

DR Pfam; PF00595; PDZ; 3.

DR Pfam; PF00018; SH3; 1.

DR ProDom; PD000066; SH3; 1.

DR SMART; SMC0072; GuKC; 1.

DR SMART; SMC0228; PDZ; 3.

DR SMART; SMC0326; SH3; 1.

DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.

DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.

DR PROSITE; PS50106; PDZ; 3.

DR PROSITE; PS50002; SH3; 1.

DR Transducer; SH3 domain; Alternative splicing; Repeat.

DOMAIN 40 126

FT	DOMAIN	154	244	PDZ 2.
FT	DOMAIN	486	566	PDZ 3.
FT	DOMAIN	600	670	SH3.
FT	DOMAIN	788	960	GUANYLATE KINASE.
SQ	SEQUENCE	960 AA;	102468 MW;	BF87A4262F1B6AD5 CRC64;

Query Match 5.4%; Score 147; DB 1; Length 960;

Best Local Similarity 21.4%; Pred No. 0.36;

Matches 123; Conservative 60; Mismatches 207; Indels 184; Gaps 26;

QY	12	DNFQSQGGQGFPAIFGQAMALAGQIRSGGSGPTVHIGTAPLGLGVVDNNGNGARVORV	71
Db	39	DIQLERNGSLGFSI-----AGGTDNPHIGTDTSIY-----TKLI	75
QY	72	VCSAPAASIGISTGVITAVDQAPINSA--TAMADALNGHFGDVISV-----TWQTK	122
Db	76	SGAAADGRSLINDIIVSVDVSVVDPHASVDALK--KAGNVVKKLVKKRGRTATTP	133
QY	123	SGGTRIGNV--TLAEGPPA--EFLVPRGSMRAFIIDPTTISAIDGLYLLGIG---IPNQG	176
Db	134	AAGSAAGDARDSAAGPKVIEDLVKGGKGLG-----SIAG-----GIGNQHIPPEN	181
QY	177	GILYSSL-----EYFEKALEELAAAFPGDGLGSAADK	209
Db	182	GIYVTKLTDGGRQVDGRSLSGIKLIAVRTNGSEKNLENVTHELAVA-----TLKSIDK	236
QY	210	Y-----AGKNRHNHVPFQSLADLDRLI-----SLIHQANAVQTTTRDILE	250
Db	237	VTLIIGKTQHLTTSASGGGGGLSSQQLSQSQSLATSQSQSVHQQHATPMVNSQST	296
QY	251	GAKKGLEFVRVAVDLTIPVVGHALSAFAFPFCAGAMA---VVGALAYLVVKTLLNA	307
Db	297	GA---LNSMGQTVDSPSPQAAAAVAAAASASASVIAASNTTNTTNTTNTTATATAS	353
QY	308	TQLLKILAKLAELVAAAIADISDVADIKIGLGEVWFITNALGLKELMKLGTGVTG	367
Db	354	NDSSKLPSLGNSSISNSNSN-----SNINININ-----389	
QY	368	LFSRQWNLSEFFAGVPLGTATGSLQVTLGFGAAGLSASSGLAHADSLASSLPALA	427
Db	390	---NNNNSSSSSTATVAAATPTAA-----SAAAAASS--PPANSFYNNASMPALP	436
QY	428	GIIGG-----GSGFGGLPSLAQV-----HAASFEQALR-PRA-----DGPVGA	463
Db	437	VESNQTNNRSQSPQPGSRVASTNVLAAPPGTPRAVSTEDITREPTITIQKPGOOL	496
QY	464	AAEQVGGGSQLVSAQSQQM-----GGPVGWG 490	
Db	497	GFNIVGGD-----GGGIYVSFILAGGPADLG 523	

RESULT 9

YI40.MYCTU	STANDARD;	PRT;	515 AA.
ID YI40.MYCTU			
AC Q50594;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Hypothetical PE-PGRS family protein Rv1840C.			
GN Rv1840C OR MT1888 OR MTCY1A11.04 OR MTCY359.33.			
OS Mycobacterium tuberculosis.			
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX NCBI_TaxID=1773;			
RN			
RC SEQUENCE FROM N.A.			
RC STRAIN=H37Rv;			
RX MEDLINE=98295987; PubMed=9634230;			
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA Gordon S.V., Bigmeier K., Gas S., Barry C.E. III, Tekai F.,			
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			

Best Local Similarity	28.4%;	Pred. No. 0.41;	
Matches	89;	Conservative	17; Mismatches 119; Indels 88; Gaps 15;
QY	257	EFVRPVAVDITYPVCHALSAFAQP-	-----FCAG-----ANAVVCGA--LAYLVV 301
Db	9	EVIAAATDIASL-	---GSSISAANAANAANTALMAAGADEVSTAALFAGHQAYQAL 65
QY	302	KTLLNA--TOLLKLLAKLAEVAAIAADIIISDVADIKIGILGEVWF-	-----ITNA 350
Db	66	SAQAQAFHAFVQALITSGGGAYAAEAAPVPLDPN-	-----EFFLANTGRPLING 118
QY	351	LNGLKELWDKLT-	-----CWVTGLSRGWSNLESFAGVPLGTATSLGSLQVTLFGAA 403
Db	119	ANAPG-	-----TGANGDGMWLTNGGAGGS-----GAAGYNGAGGNGGAGNGGAG 166
QY	404	GLSASSGLAHADSLASSASLALAGIGGSGFGGLPSLAQVHAASR-	-----QALRPADG 459
Db	167	GLTNGGAGGAGGVASS-	-----GIGSGAGGNAMLFAGAGGAGGAGGVVALTGGAGG 219
QY	460	PVGA-	-----AAEQVGGQ-----SOLVASGSGQMGPVGMGMHPSSGASKGTTTTK 505
Db	220	AGGAGGNAGLLFGAAGYGGAGGTNGSALGGAGGAGGAGGLFATGCVGGSGGAG-	-----273
QY	506	KYSEGRAAGTEDA 518	
Db	274	--SSGGAGGAGGA 284	
RESULT 11			
PGCA CANFA			
ID	PGCA CANFA	STANDARD;	PRT; 2333 AA.
AC	Q28343; Q28310;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).		
GN	AGC1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fuloop C.;		
RT	"Complete coding sequence and deduced amino acid sequence of aggrecan of canine cartilage."		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 774-833 FROM N.A.		
RC	TISSUE=Cartilage;		
RX	MEDLINE=9512852; PubMed=7827755;		
RA	Barry F.P., Neame P.J., Sasse J., Pearson D.;		
RT	"Length variation in the keratan sulfate domain of mammalian aggrecan."		
RL	Matrix Biol. 14:323-328(1994).		
RN	[3]		
RP	SEQUENCE OF 1830-2333 FROM N.A.		
RC	TISSUE=Cartilage;		
RX	MEDLINE=9335255; PubMed=8349621;		
RA	Fuloop C., Walcz E., Valyon M., Glant T.T.;		
RT	"Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module."		
RL	J. Biol. Chem. 268:17377-17383(1993).		
CC	-!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A		

CC	REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.		
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).		
CC	-!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.		
CC	-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).		
CC	-!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.		
CC	-!- SIMILARITY: Contains 4 link domains.		
CC	-!- SIMILARITY: Contains 1 C-type lectin family domain.		
CC	-!- SIMILARITY: Contains 1 Sushi (SCR) domain.		
CC	-!- SIMILARITY: Contains 1 EGF-like domain.		
CC	-!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).		
CC	EMBL; U65989; AAB06238.2; -		
DR	EMBL; S74662; AAC60527.1; -		
DR	EMBL; L07054; -; NOT_ANNOTATED_CDS.		
DR	PIR; I46998; I46998.		
DR	HSSP; P08709; IBP9.		
DR	InterPro; IPR000152; Asx hydroxyl.		
DR	InterPro; IPR000742; EGF 2.		
DR	InterPro; IPR001881; EGF Ca.		
DR	InterPro; IPR006209; EGF-like.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003006; Ig MHC.		
DR	InterPro; IPR001304; Lectin_C.		
DR	InterPro; IPR000538; Link.		
DR	InterPro; IPR003324; SGXSG.		
DR	InterPro; IPR000436; Sushi_SCR_CCP.		
DR	Pfam; PF00008; EGF; 1.		
DR	Pfam; PF00047; Ig; 1.		
DR	Pfam; PF00059; Lectin_C; 1.		
DR	Pfam; PF02339; SGXSG; 66.		
DR	Pfam; PF00084; sushi; 1.		
DR	Pfam; PF00193; Xlink; 4.		
DR	PRINTS; PR01265; LINKMODULE.		
DR	PRINTS; PR00356; ANTIFREEZEII.		
DR	ProDom; PD000318; Link; 4.		
DR	SMART; SM00032; CCP; 1.		
DR	SMART; SM00034; CLECT; 1.		
DR	SMART; SM00179; EGF_CA; 1.		
DR	SMART; SM00445; LINK; 4.		
DR	PROSITE; PS00010; ASX HYDROXYL; 1.		
DR	PROSITE; PS00615; C-TYPE LECTIN_1; 1.		
DR	PROSITE; PS00041; C-TYPE LECTIN_2; 1.		
DR	PROSITE; PS00022; EGF_1; 1.		
DR	PROSITE; PS01187; EGF_CA; 1.		
DR	PROSITE; PS00835; IG LIKE; 1.		
DR	PROSITE; PS00290; IG MHC; 1.		
DR	PROSITE; PS01241; LINK; 4.		
KW	Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;		
FT	EGF-like domain; Repeat; Immunoglobulin domain.		
FT	POTENTIAL.		
FT	SIGNAL 1 16		
FT	CHAIN 17 2333		
FT	DOMAIN 34 147		
FT	DOMAIN 170 247		
FT	DOMAIN 268 349		
FT	DOMAIN 513 590		
FT	DOMAIN 611 652		
FT	DOMAIN 2081 2117		
FT	EGF-LIKE, CALCIUM-BINDING (POTENTIAL).		

FT DOMAIN 2130 2245 C-TYPE LECTIN.
FT DOMAIN 2249 2307 SUSHI.
FT DOMAIN 248 140 GI-A.
FT DOMAIN 152 247 GI-B.
FT DOMAIN 253 349 GI-B'.
FT DOMAIN 495 589 G2-B.
FT DOMAIN 596 691 G2-B'.
FT DOMAIN 694 816 KS.
FT DOMAIN 819 1334 CS-1.
FT DOMAIN 1385 2079 CS-2.
FT DOMAIN 2080 2333 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 518 589 BY SIMILARITY.
FT DISULFID 542 583 BY SIMILARITY.
FT DISULFID 616 691 BY SIMILARITY.
FT DISULFID 640 661 BY SIMILARITY.
FT DISULFID 2085 2096 BY SIMILARITY.
FT DISULFID 2090 2105 BY SIMILARITY.
FT DISULFID 2107 2116 BY SIMILARITY.
FT DISULFID 2123 2134 BY SIMILARITY.
FT DISULFID 2151 2243 BY SIMILARITY.
FT DISULFID 2219 2235 BY SIMILARITY.
FT DISULFID 2250 2293 BY SIMILARITY.
FT DISULFID 2279 2306 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2333 AA; 240573 MW; 899ED78F3508B596 CRC64;
Query Match 5.2%; Score 143.5; DB 1; Length 2333;
Best Local Similarity 21.6%; Pred. No. 1.5;
Matches 127; Conservative 60; Mismatches 183; Indels 219; Gaps 27;
QY 41 GGSPT-----VHIGTAPLGLGVNDNGNGARVQVVGSAASLGISTGDVITAV 91
DB 706 GTTPSPVEDWPTQVGP.VPSVP-----MGEETAI 738
QY 92 DGAPI--NSATAMADALN--GHPGDVISTVMTQKSGGTRTGNVTLAEPG----- 137
DB 739 LDFTIEPENQTEWEPAYSPAGTSPLPFGIPPTWPTTSTATEES---TEGSPGTEVPVSVE 794
QY 138 ---PAEFLVPRGMSRAPIIDPTISADGLYDLGLI-----GIPNOGGILY 180
DB 795 EPSPEEPFWEELSTLPPGSPGTLPSCGSEASGVPEVSGDFTGSGEVSGHPDSSGOLS 854
QY 181 SSLEYFEKALELAAAFPGDGLWGLAADKYAKGNHNVFFQELADLDRQLISL----- 234
DB 855 G-----BSAGLPSDELDDSSGLTSVAVSG-----LASGDEDRITLSSIPKV 895
QY 235 ----THDQNAVQTRDILEGAKKLEFVRPVAVDLTYIPVGHLSAFAOPACAGAMA 290
DB 896 EGEGLTSASGVEDLSGLPSG-REGLETSTSGVGLSGLP-SGGL----- 939
QY 291 VVGALAVLWKTLLINATOLLKULAKLAELVAAAIADI-----TSDVADIKG 338
DB 940 -----EVSASGVEDLSGLPSGGEPTSTSGVGLSRL 971
QY 339 ILGEVWFETNALNGKELWKLTGVTGLPSRGNLSFAGVPLGTG-----A 389
DB 972 PSGEPEVPSASGV-----GELSGLPS-GREGLETSTSGVEDLSGLPSGGEPEAS 1019
QY 390 TSGLSQVTGL-----FGAAGLSASSGLAHADSLASSAS-LPALAGTGGG-----SG 434
DB 1020 TSGVGLSRLPSGGEPEVPSASGVEDLSGLPSGGELEASGVGLSGLPSGGEPEASAG 1079

QY 435 FGLGLPSLAQVHAASRTRQALRPADGP-VGAAAEQ-----VGGSQLVSAGSQGMGG-PV 487
DB 1080 VGLSRL-----PSGGEPEVPSASGVEDLSGLSSESPEASASGVGLSGLPS 1126
QY 488 GMGMHPSS---GASKGTTTKYSEGAAAGTDAERAP-----VEADAGG 529
DB 1127 GREGLTSASGVGLSGLPSGEGQEASGVEDLSRLPSGEGPEASAG 1175
RESULT 12
PGCA RAT
ID PGCA RAT STANDARD; PRT; 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
GN AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8087070; PubMed=3693370;
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones";
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RP REVISION TO 698.
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RP SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=8250698; PubMed=2424893;
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan";
RL J. Biol. Chem. 261:8108-8111(1986).
CC -I- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -I- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -I- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2. COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -I- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -I- SIMILARITY: Contains 4 link domains.
CC -I- SIMILARITY: Contains 1 C-type lectin family domain.
CC -I- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -I- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

Query Match

RESULT 15

PGCA MOUSE STANDARD; PRT; 2132 AA.

AC Q61282; Q64021;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 DE AGC1 OR AGC.
 GN AGC1 OR AGC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Cartilage;
 RX MEDLINE=95104847; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fuellep C., Horvath P., Doege K.J., Glant T.J.
 RA "Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan."
 RL Genomics 22:364-371(1994).
 RN [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kinata K., Line S., Strong D., Gao L.-Y., Kozak C.A., Yamada Y.
 RA "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene."
 RL Nat. Genet. 7:154-157(1994).
 RN [3]
 RP INTERACTION WITH FBLN1.
 RX PubMed=10400571;
 RA Asberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
 RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and versican."
 RL J. Biol. Chem. 274:20444-20449(1999).
 CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -!- SUBUNIT: INTERACTS WITH FBLN1.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION IN THE G1 DOMAIN).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 4 link domains.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L07049; AAC37670.1; .
 DR EMBL; S73722; AAB32160.1; .
 DR EMBL; S73721; AAB32160.1; JOINED.
 DR PIR; A55182; A55182.
 DR HSSP; P28066; 1TSG.
 DR MGI; MGI:99602; Agcl.
 DR InterPro; IPR002353; AntifreezeII.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig.MHC.
 DR InterPro; IPR003596; Ig.V.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXSG.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF02339; SGXSG; 60.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR PRINTS; PR01285; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00406; LGV; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 DR PROSITE; PS00290; IG.MHC; 1.
 DR PROSITE; PS01241; LINK; 4.
 KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 34 147 IG-LIKE V-TYPE.
 FT LINK 1.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 G1-A.
 FT DOMAIN 152 247 G1-B.
 FT DOMAIN 253 349 G1-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 597 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.
 FT DOMAIN 1917 2132 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 590 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1922 1933 BY SIMILARITY.
 FT DISULFID 1950 2042 BY SIMILARITY.
 FT DISULFID 2018 2034 BY SIMILARITY.
 FT DISULFID 2049 2092 BY SIMILARITY.
 FT DISULFID 2078 2105 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).

```
FT CARBOHYD 387 387 N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD 611 611 N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD 667 667 N-LINKED (GLCNAC . .) (POTENTIAL)
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC . .) (POTENTIAL)
FT SITE 1171 1173 CELL ATTACHMENT SITE (POTENTIAL)
SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDFC6C8DA163 CRC64;

Query Match 5.2%; Score 141; DB 1; Length 2132;
Best Local Similarity 20.6%; Pred. No. 1.9;
Matches 119; Conservative 76; Mismatches 200; Indels 184; Gaps 28;

QY 23 FAIPICQAMAIAGQIRSGSGPTVHIGPTAFGLGLVVDNNGNGARVQVVGSA-PAASL- 80
Db 799 FAVPSMTLPGSGE---ASGAPDLS-----GDFGTSGDASGRLLDSSGQPSGGIE 844
QY 81 -GISTGCV-----ITAVGAPINATAMADALNGHHPGDVISVTWQKSGGTRTGNVTL 133
Db 845 SGLPSGLDSSGLSPVSSGLPVESGSAGD---GEVWSHTPTVGRLPSSGES----- 895
QY 134 AEGPPAEFLVPRGMSRAFIIDPTISADGLYDLIGIPNOGGILYSLEYFEKALEEL 193
Db 896 -----PEGSASAS-----GTGDL--SGLPSGGEITETSTGAER----- 927
QY 194 AAAPF--GQWLGSADKYAGKRNHVNFPQSLADLDRLISLIHQANAVQTRDILEG 251
Db 928 TSGLPSSGGGLETSTSGV-----DDVSGIPTGRELET 960
QY 252 AKKGLFVRPV-----AVDLTYIFVVGHALSAAPFCACAMAVV--GGAL 296
Db 961 SASGVEDLSGLPSGEGSSETSGIEDISVLTGGSLETSAG--VGDLGLPSGGES 1017
QY 297 AYLWVKTLLNATQLLKLAKLAELVAAAIADI-----ISDVADIKIL 340
Db 1018 LETSASGAEDVTQLPTERGGL-ETSASGVEDITVLPTGRESLETSASGVEDVSLPSGRE 1076
QY 341 GEWWEFITNALNGLKELWKLGTWITGLPSRCWSNLES-----PFAGVP---GUTGATS 391
Db 1077 G-----LETSASGIEDI-----SVPTAEGLDTSASGCVYSGIPSGDGTETSA 1122
QY 392 GLSQVTGL-FGAAGLSASS-----GLAHADSLASASLPALAGIGGSG----- 434
Db 1123 GVEDVSLPSGGEGLETSASGVEDLGPSTRDSLETSASGVDTVPFPGRGDPETSVSGVG 1182
QY 435 --FGGLPSLAQVHAASTROALRPADGPVGAANEQVGGSQLVSAQSGQWG---GPVG 488
Db 1183 DDFSGLP-----SGKGLETSASGAEDLSGLPSGKEDLVGSASGALDFKLPPTLIG 1234
QY 489 MGGMHPSSGASKGTTTKYS-----EGAAAGTEDAERAP 522
Db 1235 SGQTPVNGFPFSG-FSGEYSGADIGSPSSGLPDFSGLP 1272
```

Search completed: November 21, 2003, 16:04:26
Job time : 8.35556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 15:56:56 ; Search time 26.5222 Seconds
(without alignments)
5244.295 Million cell updates/sec

Title: US-09-688-672A-64
Perfect score: 2737
Sequence: 1 MHHHHTAASDNFQLSQGG.....RAPVEADAGGQKVLVRNVV 539

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1963	71.7	392	16 Q06267	Q06267 mycobacteri
2	1263	46.1	394	16 Q49722	Q49722 mycobacteri
3	876	24.7	355	16 Q07175	Q07175 mycobacteri
4	474.5	17.3	361	2 Q03020	Q03020 mycobacteri
5	460.5	16.8	354	16 Q9CCV9	Q9CCV9 mycobacteri
6	385.5	14.1	402	16 P96213	P96213 mycobacteri
7	200.5	7.3	446	16 Q8VKA4	Q8VKA4 mycobacteri
8	199.5	7.3	464	16 Q53896	Q53896 mycobacteri
9	188.5	6.9	382	16 Q9CD67	Q9CD67 mycobacteri
10	188.5	6.9	452	2 Q325G6	Q325G6 mycobacteri
11	184.5	6.7	459	16 Q8XQX4	Q8XQX4 raiistonia s
12	180.5	6.6	1468	5 Q9GUB5	Q9GUB5 galleria me
13	169	6.2	496	16 Q53631	Q53631 mycobacteri
14	169	6.2	533	16 Q8VKR5	Q8VKR5 mycobacteri
15	168	6.1	1417	16 Q9HVN6	Q9HVN6 pseudomonas
16	167	6.1	916	9 Q80116	Q80116 bacterioph

ALIGNMENTS

RESULT 1

006267
ID O06267 PRELIMINARY; PRT; 392 AA.
AC O06267;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein RV3616C.
GN RV3616C OR MT3718 OR MTCY07H7B.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

17 166.5 6.1 1018 16 Q9HWU6
18 163 6.0 536 16 Q8FND6
19 163 6.0 911 16 Q8DXK6
20 162.5 5.9 1460 16 Q8K6B1
21 161.5 5.9 542 16 Q9FBK9
22 159.5 5.8 810 11 Q8ES29
23 159.5 5.8 860 11 Q8C9L8
24 159 5.8 1467 16 Q8RY75
25 159 5.8 3961 16 Q8P942
26 158 5.8 667 16 Q8XRN8
27 157.5 5.8 683 16 Q8G791
28 154 5.6 519 16 Q93J30
29 154 5.6 1329 16 Q06810
30 153.5 5.6 1488 16 Q8NPKO
31 153 5.6 2751 16 Q8XUKO
32 152.5 5.6 757 4 Q14234
33 152.5 5.6 1610 16 Q92KQ8
34 152 5.6 409 16 P95182
35 152 5.6 628 16 Q8VJ19
36 151.5 5.5 375 16 Q8DGB7
37 151.5 5.5 940 16 Q8K755
38 151.5 5.5 3145 16 Q98MG7
39 151.5 5.5 3659 16 Q98LNG
40 151 5.5 525 16 Q8PMV4
41 151 5.5 588 16 Q50396
42 151 5.5 1408 16 Q8VK17
43 150.5 5.5 937 9 Q9G097
44 150.5 5.5 2183 16 Q8PKH6
45 150 5.5 436 16 Q8VIX6

Q9HWU6 pseudomonas
Q8FND6 corynebacte
Q8DXK6 streptococc
Q8K6B1 streptococc
Q9FBK9 streptomyc
Q8ES29 mus musculu
Q8C9L8 mus musculu
Q8RY75 deinoococcus
Q8P942 xanthomonas
Q8XRN8 raiistonia s
Q8G791 bifidobacte
Q93J30 streptomyc
Q06810 mycobacteri
Q8NPKO corynebacte
Q8XUKO raiistonia s
Q14234 homo sapien
Q92KQ8 rhizobium m
P95182 mycobacteri
Q8VJ19 mycobacteri
Q8DGB7 synchococc
Q8K755 bacillus ha
Q98MG7 rhizobium l
Q98LNG rhizobium l
Q8PMV4 xanthomonas
Q50396 mycobacteri
Q8VK17 mycobacteri
Q9G097 lactococcus
Q8PKH6 xanthomonas
Q8VIX6 mycobacteri

```

DR ENBL; Z95557; CAB08950.1; -.
DR ENBL; AB007171; AAK48077.1; -.
DR TIGR; MT3718; -.
DR TubercuList; RV3616c; -.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 192 192
SQ SEQUENCE 392 AA; 39888 MW; 82BBA8DD9D6F567 CRC64;

Query Match      71.7%; Score 1963; DB 16; Length 392;
Best Local Similarity 99.7%; Pred. No. 2.3e-102;
Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 208 DKYAGKRNHNHVFQELADRLQLSLIHQDQNAVQTTDRDILEGAKKGLFVRPVAVDLT 267
DB 61 DKYAGKRNHNHVFQELADRLQLSLIHQDQNAVQTTDRDILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAAFPQAFPCAGAMVVGALAYLVVKTLLNATOLLKLLAKLAEVAAIA 327
DB 121 YIPVVGHALSAAFPQAFPCAGAMVVGALAYLVVKTLLNATOLLKLLAKLAEVAAIA 180

QY 328 IISDVADIIGILGEVWEFITNALNGELKELWDLKLTGWVTGLFSRGSNLESPFAGVPG 387
DB 181 IISDVADIIGILGEVWEFITNALNGELKELWDLKLTGWVTGLFSRGSNLESPFAGVPG 240

QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGTGGSGFGGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGTGGSGFGGLPSLAQVHAA 300

QY 508 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 539
DB 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 2
Q49722 ID Q49722 PRELIMINARY; PRT; 394 AA.
AC Q49722;
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ML0405.
GN ML0405 OR B1620_C2_213 OR MCL383.01.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R., Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D., Mungall K., Basham D., Brown D., Chillingworth T., Connor R., Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N., Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S., Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R., Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

(3)
RN SEQUENCE OF 231-394 FROM N.A.
RP Oliver K., Harris D., Parkhill J., Barrall B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
(4)
RN SEQUENCE OF 231-394 FROM N.A.
RP MEDLINE=93188700; PubMed=8446027;
RX Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae.";
RL Mol. Microbiol. 7:197-206(1993).
DR ENBL; AL583918; CAC29913.1; -.
DR ENBL; U00015; RAC43223.1; -.
DR ENBL; Z97179; CAB03940.1; -.
DR Leproma; ML0405; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 279 299 POTENTIAL.
SQ SEQUENCE 394 AA; 40754 MW; D0E455C74ED7A10C CRC64;

Query Match      46.1%; Score 1263; DB 16; Length 394;
Best Local Similarity 82.3%; Pred. No. 4e-63;
Matches 248; Conservative 55; Mismatches 89; Indels 2; Gaps 1;

QY 148 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAFPGDGLGSA 207
DB 1 MSRAFIIDPTLKAELAWHALLGIGVNDGGVLYSSLSFFEKALEHLAAAFPGDGLGSA 60

QY 208 DKYAGKRNHNHVFQELADRLQLSLIHQDQNAVQTTDRDILEGAKKGLFVRPVAVDLT 267
DB 61 DKYAGKRNHNFQELADRLQLSLIHQDQNAVQTTDRDILEGAKKGLFVRPVAVDLT 120

QY 268 YIPVVGHALSAAFPQAFPCAGAMVVGALAYLVVKTLLNATOLLKLLAKLAEVAAIA 327
DB 121 YIPVVGHALSAAFPQAFPCAGAMVVGALAYLVVKTLLNATOLLKLLAKLAEVAAIA 180

QY 328 IISDVADIIGILGEVWEFITNALNGELKELWDLKLTGWVTGLFSRGSNLESPFAGVPG 387
DB 181 WSDGVAILKIGIVDHLWHFIAGALTGLKDIIVEKIIHWFGLFSHWSRLHSPFGGIPGLS 240

QY 388 GATSGLSQVTLGFGAAGLSAGSLAHADSLASASLPALAGTGGSGFGGLPSLAQVHAA 447
DB 241 GATSGLSQVTLGFGVPLAGSGLSGLSTENLPSLAGVAGLGLSLPQLAQHAA 300

QY 448 STRQALRPRADGPVGAAGAEQVGSQSLVSAQSQSGMGVGMGHPSPSSGASK--GTTTK 505
DB 301 STRQTRSQAGVSAELSTEQFGGQBPVSAQSQSGMGVGMGHPSPSSGASK 360

QY 506 KYSEGAAGTDAERAPVEADAGGQKVLVRNVV 539
DB 361 KYSEGAAGTDAERAPIEVQSGGKRALAQHV 394

RESULT 3
Q07175 ID Q07175 PRELIMINARY; PRT; 355 AA.
AC Q07175;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 34.9 kDa protein (Serine protease, putative).
GN p5EA OR RV0125 OR MTCI4188.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

```

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.",
 RL Nature 393:537-544 (1998).
 RN (2).
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.",
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: Z96071; CAB09453.1; -;
 DR EMBL: AE006925; AAK44357.1; -;
 DR TIGR: MT0133; -;
 DR TubercuList; RV0125; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_proteas_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydroxylase; Hypothetical protein; Serine protease; Protease;
 KW Complete proteome.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 24.7%; Score 676; DB 16; Length 355;
 Best Local Similarity 99.2%; Pred. No. 2.8e-30;
 Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 7 HTAASDNFQLSQGGQGAIPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGA 66
 Db 223 NTAASDNFQLSQGGQGAIPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGA 282
 Qy 67 RVQRVVGSAAPASLGISTGVDITAVDGPAINSATAMADALNGHPGVDIVSVTWQKSGGT 126
 Db 283 RVQRVVGSAAPASLGISTGVDITAVDGPAINSATAMADALNGHPGVDIVSVTWQKSGGT 342
 Qy 127 RTGNVTLAEGPPA 139
 Db 343 RTGNVTLAEGPPA 355

RESULT 4
 ID Q50320 PRELIMINARY; PRT; 361 AA.
 AC Q50320;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 34Kda protein precursor.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JD88/107;
 EX MEDLINE=95005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease

RT expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982(1994).
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL: Z23092; CA80638.1; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydroxylase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL.
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 17.3%; Score 474.5; DB 2; Length 361;
 Best Local Similarity 69.7%; Pred. No. 5.6e-19;
 Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

Qy 8 TAASDNFQLSQGGQGAIPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 67
 Db 231 TAATDSYKMS-GGQGAIPIGMAVANQIRSGAGSNVHIGPTAFGLGVVDNNGNGAR 289
 Qy 68 VQRVVGSAAPASLGISTGVDITAVDGPAINSATAMADALNGHPGVDIVSVTWQKSGGT 127
 Db 290 VQRVVTGPAAGAAAGIAPGDVITGVTVPINGATSVTEVLVPHHPGDTIAVHRSVDGGER 349
 Qy 128 TGNVTLAEGPPA 139
 Db 350 TANITLAEQPPA 361

RESULT 5
 Q9CCY9 PRELIMINARY; PRT; 354 AA.
 ID Q9CCY9
 AC Q9CCY9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Probable secreted serine protease.
 GN M2659.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.",
 RL Nature 409:1007-1011(2001)
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL589326; CAC32191.1; -;
 DR Leproma; M2659; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.

```
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC92A4A CRC64;

Query Match 16.8%; Score 460.5; DB 16; Length 354;
Best Local Similarity 66.2%; Pred. No. 3.3e-18;
Matches 88; Conservative 18; Mismatches 26; Indels 1; Gaps 1;

QY 7 HTAASNFOQLSQGGGQFAIPGQMAIAQIRSGGSPTHVIGPTAFGLGVVDNNGGA 66
DQ 223 NTAATDNYKM-LGGQGFAPIGQMEVWGAIRSGAGSNTVHIGPTAFGLGVLDNNGGA 281
QY 67 RVORVVGSAAPAGSLGSTDGVTAVDAPINSATAMADALNGHPCDVISVTWQKSGT 126
DQ 282 RVARVVATGPAVAGLSVGDITSDVGVFISEATATNVLVPHHPGETVAVNRSAGGDD 341
QY 127 RTGNVTLAEGPPA 139
DQ 342 LTANVTLAEGPPA 354

RESULT 6
P96213 PRELIMINARY; PRT; 402 AA.
AC P96213;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical protein Kv3864.
GN RV3864 OR MTCY01A6.04C OR MT3978.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z83864; CAB06237.1; -.
DR EMBL; AF5007189; AAK48347.1; -.
DR TIGR; MT3978; -.
DR TubercuList; Kv3864; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 402 AA; 42068 MW; B400E0E22D482765 CRC64;

Query Match 14.1%; Score 385.5; DB 16; Length 402;
Best Local Similarity 30.4%; Pred. No. 6.3e-14;
Matches 130; Conservative 47; Mismatches 156; Indels 95; Gaps 13;

QY 157 TTSADIGLYDLGIGIPNOGGILYSLEYFEKALEELAAAFPGDWLGSAAADKYAGKNRN 216
DQ 8 TAASDNFQLSQG--GGFAIPGQMAIAQIRSGGSPTHVIGPTAFGLGV--DNNGN 64
QY 8 TAASDNFQLSQG--GGFAIPGQMAIAQIRSGGSPTHVIGPTAFGLGV--DNNGN 64
```

```

Db 319 TLGADSADAQSGSIGLGFPAIPVDQAKRIADELISTGKA-----SHASLGQVQVINDKDT 372
QY 65 GARVQVVGAPASLSIGSTGVDITAVDGPINSATAMADALNGHHPCDVISVTWQTKSG 124
D 373 GAKIVEVVGAAANAGVPGVGVVTKVDDRPINSADALVAARSKAPGATVALTFQDPDSG 432
QY 125 GTRTGNVTLAE 135
D 433 GSRTVQVTLGK 443

RESULT 8
ID O53896 PRELIMINARY; PRT; 464 AA.
AC O53896;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative serine protease.
GN RV0983 OR MTW044.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh J., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL021999; CAA17582.1; -.
DR TubercuList; RV0983; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 464 AA; 46452 MW; AE93BFCC3E1EC8F CRC64;

Query Match 7.3%; Score 199.5; DB 16; Length 464;
Best Local Similarity 38.9%; Pred. No. 0.002;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 8 TAASNFQLSQ--CGGPAIPGQAMAIAGQIRSGGSPVTHGPTAFGLGLGVV-DNNGN 64
D 337 TLGADSADAQSGSIGLGFPAIPVDQAKRIADELISTGKA-----SHASLGQVQVINDKDTL 390
QY 65 GARVQVVGAPASLSIGSTGVDITAVDGPINSATAMADALNGHHPCDVISVTWQTKSG 124
D 391 GAKIVEVVGAAANAGVPGVGVVTKVDDRPINSADALVAARSKAPGATVALTFQDPDSG 450

QY 125 GTRTGNVTLAE 135
D 451 GSRTVQVTLGK 461

RESULT 9
Q9CD67 PRELIMINARY; PRT; 382 AA.
ID Q9CD67

```

```

Q9CD67;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL58917; CAC29684.1; -.
DR Leproma; ML0176; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3DD8DD8AE32A80D CRC64;

Query Match 6.9%; Score 188.5; DB 16; Length 382;
Best Local Similarity 39.8%; Pred. No. 0.0065;
Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 9 AASDNFQSGGQGFPAIPGQAMAIAGQIRSGGSPVTHGPTAFGLGLVVDNNGN-GAR 67
D 258 ADSGDAQSGSIGLGFPAIPVDQAKRIADELISTG--KATH----ASLGQVATDKGTPGAK 311
QY 68 VQRVWGSPAASLSIGSTGVDITAVDGPINSATAMADALNGHHPCDVISVTWQTKSGGTR 127
D 312 VMDWAGGAANAANAVKGVVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDSGSSR 371

QY 128 TGNVTLAE 135
D 372 TVQVTLGK 379

RESULT 10
Q9Z5G6 PRELIMINARY; PRT; 452 AA.
ID Q9Z5G6
AC Q9Z5G6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative serine protease.
GN MLCB373-28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Raftery M.A.,
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3].
RN SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Englemer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RL of Mycobacterium leprae.";
RM Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00899; trypsin; 1.
DR SMART; SMC0228; PDZ; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 452 AA; 45130 MW; 6CA675BE0911F983 CRC64;

Query Match 6.9%; Score 188.5; DB 2; Length 452;
Best Local Similarity 39.8%; Pred. No. 0.0081;
Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

QY 9 AASDNFQLSGGOGGPAIPGQMAIAQIRSGGSGPTVHIGTAFILGLGVVDNNGN-GAR 67
DB 328 ADSGDAQSGSIGLGFPAIPVDQAKRIADELISTG--KATH---ASLGVQVATDKGTGPAK 381
QY 68 VQRWGSAPAAASIGISTGVITAVDCAPINSATAMADALNGHHPGDVISTVWTKSGGTR 127
DB 382 VMDVAGGAANAAPKGVLLKYVDRLLISSADALVAARSKAPGDKVSLTQDQSGSSR 441
QY 128 TGNVTIAE 135
DB 442 TVQVTLGK 449

RESULT 11
Q8XQX4 PRELIMINARY; PRT; 459 AA.
ID Q8XQX4
AC Q8XQX4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative hemagglutinin-related transmembrane protein.
GN RSP1094 OR RS02600.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]_TaxID=305;
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646082; CAD18245.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 459 AA; 41117 MW; FE06C4144483ACC9 CRC64;

Query Match 6.7%; Score 184.5; DB 16; Length 459;
Best Local Similarity 24.7%; Pred. No. 0.014;
Matches 133; Conservative 60; Mismatches 200; Indels 145; Gaps 28;

QY 32 AIAGQIRSGG---GSPTVHIGTAFILGLGVVDNNGNARVQVVGSAAPASLIGISTGDV 87

```

```

13 AVAGLLILGGCASSGS-----GDTSGT-LGSSSSGNGA--SGSAGGSSGSDGTSAGTS 64
QY 88 ITAVDGAPINSATAMADALNGHHPGDVISTVWTKSG-GTRTGNVTIAEGPPAEFLVPRG 146
DB 65 GTNGSGTTPTLTA-GTVLN--NSGGVVTAGTSLGIGTSLKNANLP-----LVPS 114
QY 147 SMSRAFIIDPTISAIDGLYLIGIPNQGILYSSLEVEFEKALEELAAAPPGDWLGS 206
DB 115 AQ-----SGLGGVVENLGAIVSALG-----AGVDG-LGS- 143
QY 207 ADYAGKRNHNVFFOELADRLQLISLIHQANAVOTTRDILEGAKKG-LEFVRPVAVD 265
DB 144 ---IGSNPNPIG--TTVASTG---NVVTGVTNTSAGTLVGLGTGQLSLAPVTT 193
QY 266 LTYIPVVGHALSAAFQAFPCAGAMVVGAL--AYLVVKTLINATOLLKLAELVAA 323
DB 194 L-----AGALTQVQAVTNAGTTLGTTLSTGTPVEQTLQSLTAIVP 234
QY 324 AIADIISDVADIKILGEVMEFITNALNGKLWKLTGVTGLFSGRWSNLESFFAGV 383
DB 235 LTSIIIGTQTV-----GATTCGWTGTPANSLTGLGAIANGSTLTN--ANV 279
QY 384 PGLTGATSGLSOVLGTPCAAGLSASS--GLAHADSLASSASL-PALAGICGSGGFLGPS 440
DB 280 P-----VVSNGVGVVAVGNTVASLGGVAYSASSASPLAPVTGALGGGNPLGAVTS 332
QY 441 LAQVHAASTRQALRPADGFPVGAAGVGG-----QSOLVSAQSGQMGQGPV-- 487
DB 333 VLGLGSGTSG-SPLA--PVTGVLSSVGATGSGSPLAAVTSLGSLGAAGSGSPLAP 399
QY 488 ---GMGMHRESSGASKGTTTKYSEGAACGTBA-----ERAPVADAGG 529
DB 390 VTSALGSGVGGTGG--GTLAPVASITAPVGTVATVGTSTLTGATTPAPVTGSAGG 445

RESULT 12
Q9GUB5 PRELIMINARY; PRT; 1468 AA.
ID Q9GUB5
AC Q9GUB5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Heavy-chain fibroin (Fragment).
GN FB-H.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea;
OC Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]_TaxID=7137;
RP SEQUENCE FROM N.A.
RC TISSUE=Posterior silk gland;
RA Zurovec M., Kodrik D., Yang C., Sehnael F.;
RT "Heavy-chain fibroin of Galleria mellonella L.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095239; AAG10393.1; -.
FT NON TER 1468 1468
SQ SEQUENCE 1468 AA; 122705 MW; 2DA59E1181B3DDF CRC64;

Query Match 6.6%; Score 180.5; DB 5; Length 1468;
Best Local Similarity 25.1%; Pred. No. 0.11;
Matches 138; Conservative 42; Mismatches 219; Indels 151; Gaps 22;

QY 28 GQAMATAGQIRSGGSPVHIGTAFILGLGVVDNNGNARVQVVGSAAPASLIGISTGDV 87
DB 473 GVSTAGSLGGUGGAGVSAVGP-AGAGLGGVAGGSGS-----SASAAARSAPVP 526
QY 88 ITAVDGAPINSATAMADA-----LNHHHPGDVISTVWTKSGGTRTG---- 129
DB 527 IVIEDGSSAASAAAAGSAGSLGGLGAWGPLGGIGPNEVSSAS-ATGSAAGSTGAGLG 585
QY 130 -----NVTLAGPPAEFLVPRGMSRAFIIDPTISAIDGLYDLIGI---PNQ 175

```

```
Db 586 GSGAAGSAAAGAAAGPAPVIVIEDGSSAAS-----AAAAGSAGSLGGLGAGWGTL 640
Qy 176 GGILYSLEYEKEALEELAAAFPGDGMGSAADRYAGKNRHNHNFQELADLPQLISLI 235
Db 641 GGIGPNQVS--SASATGSAAGSISGSLGSGA--AGSS-----674
Qy 236 HDQANAVOTTRDILEGAKGLLEFVRPVAVDLTYIPVUGHALSAAFOAPFCAGAAVVG-- 293
Db 675 -----LASAASGAAGAAPVI-----VIEDGSSAASAAAAGSAGSGVGLG 714
Qy 294 -GALAYLVVKTINATQLLAKLAELVAAAADIISDVADI--SASGLAHADSLASSASLPA-----LAGIGGSGFGGLPGLPSLAQV 444
Db 715 LGALGCL-----GGIGPISASSAGSAGAGLGGVGAAGTSGLG 751
Qy 353 GKELWMDKLTGWTLFSGRGSNLESFPAGVPL-TGATGSLGSLVTLFGAAG-----L 405
Db 752 GI-----GCV-----GASTAGSAGAGLGGIGAGSGSSAASAASGASGAGEVIVI 797
Qy 406 SASGLAHADSLASSASLPAAGLGGSGFGGLPGLPSLAQVHAASRQALRPADGPGVAAA 465
Db 798 DRSAAASAAAAGSAGSGFGGLGVLGWPGLGGIGPISASSASAGL-----GCVGAAG 852
Qy 466 EQ-----VGGQSOLVAQSQGMG-GPVGMMGMHPSFGSKGTTTTKYSAGAAAAGTDAER 520
Db 853 TSGLGGLGGAG--ASAAGSAGAGLGGIGAGSGSSGSSVASAASGTS--GAGEVIVIDDRSS 908
Qy 521 APVEADAGG 530
Db 909 AASAAAAGSG 918

RESULT 13
OS3631 PRELIMINARY; PRT; 496 AA.
ID OS3631
AC OS3631
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE PGSR-family protein.
GN RV0109 OR MT031.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kroch A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL Nature 393:537-544(1998).
EMBL AL021926; CAAL7303.1;
DR TubercuList; rv0109;
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Complete proteome.
SQ SEQUENCE 496 AA; 42289 MW; A3FLAD215EC11786 CRC64;

Query Match 6.2%; Score 169; DB 16; Length 496;
Best Local Similarity 25.5%; Pred. No. 0.11;
Matches 83; Conservative 26; Mismatches 107; Indels 110; Gaps 14;

Qy 262 VAVDLTYIPVGHALS---AFAQAPFCAGAAVVG-----293
```

```
Db 11 VAAATHLAGISALSTANAAAAP--TTALSVAGADEVSVLIAALFEAYAEYQALSQA 68
Qy 294 -----GALAYLVVKTINATQLLAKLAELVAAAADIISVDADI--336
Db 69 ALAFHDQFVQALNMGAVCYAAAE-ANATPEL-----QALQTVQNVLTIVNAP 115
Qy 337 -KGILGEVWEFTINAGLKLWDLKLTGWVTG-----LFSRGRHSNLESFPAGVPLGTGATS 391
Db 116 TQALLGR--PIINGANGLPN-----TGQDGGPGLLFNGGN-----GSG 155
Qy 392 GLSQVTGLFGAAGL--SASSGLAHADSLASSASLPA-----LAGIGGSGFGGLPGLPSLAQV 444
Db 156 GVDQAGNGGAAGLIGNGGSGVGGPGIAGSAGGAGAGLFGNGGPGGAGGIGITGD- 214
Qy 445 HAASTQALRPADGPGVGAAGVQVQSQSOLVAQSQGMGPGVGVGCMHPSSGASKGTTT 504
Db 215 -----GGFGAGGNAIG-----LFGSGGTGGMGGVGGMGVGNAGNGGTA 257
Qy 505 KYSEGAAGAAGTDAERAPVEADAGG 530
Db 258 GLFGHGAGAGGAGISADGGLGGGG 283

RESULT 14
Q8VKR5 PRELIMINARY; PRT; 533 AA.
ID Q8VKR5
AC Q8VKR5
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE PE PGSR family protein.
GN MT0118.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006923; AAK4341.1;
DR TIGR; MT0118;
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 533 AA; 46785 MW; 56F0115F01F4D4E5 CRC64;

Query Match 6.2%; Score 169; DB 16; Length 533;
Best Local Similarity 26.2%; Pred. No. 0.13;
Matches 89; Conservative 25; Mismatches 134; Indels 92; Gaps 15;

Qy 229 RQLISLHDAQANAVQTRDILEGAKGLLEFVR-PVAVDLTYIPVVG-----273
Db 35 RSLMSLITSPATVAAAATHLAGISALSTANAAAAPTTALSVAGADEVSVLIAALFEA 94
Qy 274 -----HALSA---AFAQAPFCAGAAVVGALAYLVVKTINATQLLAKLAELVAAA 325
Db 95 YAQEYQALSQAALAFHDQF---VQALNMGAVCYAAAE-ANATPEL-----QAL 138
Qy 326 ADIISVDADI-----KGILGEVWEFTINAGLKLWDLKLTGWVTG-----LFSRGRHSNLE 377
Db 139 QTVQNVLTIVNAPTALLGR--PIINGANGLPN-----TGQDGGPGLLFNGGN--- 188
```

```
QY 378 SFAGVPLTGATSGLSQVLTGLEAAGL--SASSGLAHADSLASSASLPA-----LAGIG 430
Db 189 -----GGGGVDQAGNGGAAAGLITNGGSGGVGGPGIAGSAGGAGGGLLFNG 238
QY 431 GSGGFGGLPLSLAQVHAASRQALRPRADGVPVGAAGQVGGQSOLVSAOCSQGGGPGVNG 490
Db 239 GPGAGGIGITGD-----GGPGAGGNAIG-----LFGSGGTGGGPGVNG 280
QY 491 GMPSSGASGTTTKYSEGAAGTDAERAPVEADAGG 530
Db 281 GVGNGGNAGNGGTAGLFGHGAGGAGGAGGAGGAGGGLGGGG 320

RESULT 15
Q9HVN6 PRELIMINARY; PRT; 1417 AA.
AC Q9HVN6
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein PA4541.
GN PA4541.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10994043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RL Nature 406:959-964 (2000).
DR EMBL: AE004867; AGO7929.1; .
DR HSSP: P22629; 1SWC.
DR InterPro: IPR006162; Ppantne attach.
DR PROSITE: PS00012; PHOSPHOPANTHEINE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1417 AA; 139958 MW; 212C916D5A55C39D CRC64;

Query Match 6.1%; Score 168; DB 16; Length 1417;
Best Local Similarity 22.1%; Pred.No. 0.55;
Matches 139; Conservative 77; Mismatches 206; Indels 206; Gaps 31;

QY 19 GGQGAIPIQQVMAIAG-----QIRSGG 41
Db 782 GGNSSFLSICNASAFGCTFDGLGNTIDNLAVYGTGAYSLFVNRGTLNRLNLERISADG 841
QY 42 GSPT---VHIGPTAFLGLGVNDN-NGNGARVORVVGSAPARSLGISTGDVITAVDGPIN 97
Db 842 AQTHNVVQVSLAAVNLGRINDVNASDIRI-----AAASKNSLGG--LVALNLGSD 893
QY 98 SATAMADALNGHHPGDVI SVTWQTSKGGTRTGNVTIAEG-----PPAEFLVP----- 144
Db 894 NASASGTLVGNRH-----TVAL--GGLAENISTARGVASISNSRADFAISGQKDKHA 944
QY 145 -----RGSMSRAFIIDPTISADIGLDLIGIGIPNQGIL-YSSLEYFEKALE 191
Db 945 SHYGAGGLVGRNRGGLIRS-----SGSQGTLSLSGHGM-NLGLVGYS-----AGGLA 992
QY 192 ELAAA--FPGDGWLGSAAQYAKNKNHNVFFQELADLDROLISLIHQANAV--QTTED 247
Db 993 DVSASVDVSGNGRG-----LYGGLICLVN-----SGIAHATASKVGTDAE 1036
QY 248 ILEGAKKLEFVRPNAVDITYPVVGHALSAAQAPFCAGAMAV-----VGGALAY----- 298
Db 1037 ALGG-----LIGNLNAAINNAAHNSAHGSDVSLQAGRYLGLGLIGNQAG 1077
```

```
QY 299 -----LVVKTLLINATQLLLKLLAKLAEVLAAAIADIISDVADIKIGIL----- 340
Db 1078 NLNVSTSGNLSGGSLIQAGGLIGLNANASLVNASAKGNVATRGAEAVGGLIGENLYGSV 1137
QY 341 -----GEVWEFITNALNGLKELWDKLTGWVTGLFSEGMNLSFFPAGVPLGTGATSGLS 394
Db 1138 INGSASGEVTDGSGKTLLGL--IGSNLGGNHSNLSKAGWVN-----AGAN--S 1181
QY 395 QVTGLFCAAGLSASSGLAHADSLASSASLPAAG--ICGGSGFGGLPGLSLAQVHAASRQA 452
Db 1182 DVGGLIG-----HNRGNGHS-TLAASGNVTGGKSRVGLGVGYNDAASLTVNASGNVSA 1235
QY 453 LRPRADGVPVGAAGQVGGQSOLVSAQG-----SQNGGPGVNG-----GMHSSGASKG 501
Db 1236 SGSEAIG--GLIGSDLRGSLMLASSHGIVNDKTSNHLGGLVGRGENTSIRSAKASGAVSG 1293
QY 502 TTTTKYSEGAAGTDAERAPVEADAGG 529
Db 1294 GAGIR-AGGLVGSLEGWQALILGASAGG 1320
```

Search completed: November 21, 2003, 16:08:18
Job time : 27.5222 secs